Learnability can be Undecidable

Travail Encadré de Recherche

Adrien Ragot
Professeurs Encadrants : Myriam Quatrini et Pierre Pudlo

17 juin 2019

Table des matières

1 Relating Compression and Learnability 5
1.1 Monotone Compression Schemes, Boosting, équivalence compression forte et faible 5
1.1.1 Definition et Objectif . . . 5
1.1.2 Une condition suffisante de compressibilité avec le choix . 8
1.1.3 Theoreme de Boosting, Decompresseur . . . 10
1.1.4 Compression faible et compression forte . . . 13
1.2 EMX Learner . . . 14
1.2.1 Definition of Learner . . . 14
1.2.2 Extension du domaine d'une application . . . 15
1.2.3 L'apprenabilité faible entraîne la compressibilité faible . . 17
1.3 How compression implies learnability . . . 18

2 Relating Compression and Cardinality 20
2.1 La compressibilité entraîne une majoration de la cardinalité . . . 20
2.1.1 Existence de points irrécouvrables . . . 20
2.1.2 Construction de scheme à compression plus forte, sur une partie stricte 22
2.1.3 Un ensemble de cardinal trop grand ne peut pas être compresser 23
2.2 Un ensemble de cardinalité suffisamment petite, est compressible 25
2.2.1 Cas initial, cas d'un ensemble au plus dénombrable. . . . 25
2.2.2 Stabilité par passage au successeur . . . 26
2.2.3 Synthèse . . . 28

3 Undecidability Of The EMX learnability 29
3.1 Resultats classiques de cardinalité . . . 29
3.2 Formule indépendante, ZFC, et déduction . . . 30
3.2.1 Formule indépendante, Axiomes de ZFC . . . 30
3.3 Indécidabilité d'un probleme d'apprentissage EMX . . . 31
3.3.1 Modèles de ZFC, Les modèles de Gödel et Cohen . . . 31
3.3.2 Le résultat d'indécidabilité . . . 33

A Mesure et Probabilité 36
A.1 Théorie de la mesure . . . 36
A.2 Loi de probabilité . . . 36
A.3 Suite de variables aléatoires . . . 37

B Bon ordre, Ordinaux et cardinaux 38
B.1 Bon Ordre et Segment Initial . . . 38
B.1.1 Bon Ordre . . . 38
B.1.2 Segment Initial . . . 38
B.2 Ordinaux, Classe Cardinale, Cardinal . . . 41
B.2.1 Ordinaux . . . 41
B.2.2 Subpotence, Equipotence . . . 42
B.2.3 Classe Cardinale . . . 42
B.2.4 Cardinaux . . . 43
B.2.5 Type D'ordre . . . 45
B.3 Alephs et Omegas . . . 48

C Théorie des Modèles **49**
C.1 Syntaxe 49
C.2 Semantique, Structure 50
C.3 Théorie, Modèle d'une théorie 51
C.4 Sous structure 51

Références **52**

Introduction

L'objet de mon travail a été d'essayer de comprendre l'article "Learnability can be undecidable" [1], c'est à dire s'approprier les notions, les objets qui interviennent dans l'article et aussi chercher à saisir, expliciter les preuves de l'article. Pour se faire j'ai donc également regarder une autre version de l'article "On a learning problem that is independent of the set theory ZFC axioms" [2], qui parfois explicite certain points de l'article. J'ai du aussi me familiariser avec certains domaines des mathématiques, nécessaires a la comprehension de l'article, notamment la théorie des probabilités, la théorie des modèles, mais aussi pour pourvoir parler formellement de cardinalité, la théorie des ordinaux (traité dans l'annexe). Aussi on cherche à avoir une vision plus globale du problème à comprendre pourquoi on se pose de tels question pourquoi parle t-on de compression, et de dimension dans ce contexte d'apprentissage, qu'est ce qui amènent les auteurs de l'article à traiter de ces notions de compression et dimension. Pour essayer de saisir ce contexte générale j'ai fait appel, entres autres, aux articles [6], [7] et [14].

Avant tout, l'article se pose des questions sur le domaine de l'apprentissage automatique théorique, il y a différentes approches à l'apprentissage, l'un d'elle est le modèle d'apprentissage PAC (pour Probably Approximately Correct) due à Leslie Valiant. L'apprentissage PAC a fait l'objet de plusieurs études, et il existe des résultats fondamentaux qui caractérise cet apprentissage. L'une de ces caractérisations est la dimension VC (pour Vapnik-Chervonenkis), une classe $\mathcal{C}$ de concept est PAC apprenable si et seulement si sa dimension VC est finie, voir par exemple le theoreme 2.1. dans [14]. L'intêret remarquable de cette caractérisation c'est qu'elle permet de passer d'un problème où il faut montrer l'existence d'une certaine fonction (un appreneur), à un problème de combinatoire. Il y a également une seconde caractérisation, par la notion de Basic Compression Scheme, une classe de concept est PAC apprenable si et seulement si, elle est compressible par un scheme de compression basique. L'article [6] montre comment la compression entraine l'apprenabilité, et plus récemment [7] montre l'implication inverse.

Au vue de ces caractérisations, l'article est motivé par le questionnement suivant ; peut on généraliser les résultats de l'apprentissage PAC à d'autres types d'apprentissages ? Et donc ici en particulier l'article traite de l'apprentissage EMX (Estimating the Maximum).

En ce sens les auteurs se demandent si il existe une notion de compression qui caractérise cet apprentissage. C'est l'objet de la première partie "Relating Compression and Learnability", dans laquelle est définit la notion de scheme de compression monotone, cette notion de scheme, différente du basic compression scheme de Littlestone et Warmuth [6], permet en effet de caractériser l'apprentissage EMX.

Dans un second temps, l'article se demande si il existe de facon similaire une notion de dimension qui caractérise l'apprentissage EMX. Il s'avère que non ; c'est le résultat plus ou moins final de l'article "Learnability can be undecidable", voir [1]. En fait, l'article montre que l'apprenabilité EMX d'une certaine classe de concept $\mathcal{F} = \mathcal{P}_{fin}([0,1])$ [1] n'est pas décidable. En fait, l'équivalence entre l'apprenabilité EMX et la compressibilité, entraîne le fait que l'apprenabi-

1. L'ensemble des parties finies de [0,1].

lité de $\mathcal{P}_{fin}([0,1])$ dépends de la cardinalité du continu. C'est cette dépendance qui entraînera l'indécidabilité. Par la suite, ce résultat d'indécidabilité entraîne l'impossibilité d'une notion de dimension qui caractériserait l'apprentissage EMX.

Dans ce document on suivra le plan suivant ; dans une première partie, on lie les notions d'apprenabilité EMX et de compressibilité. Dans la seconde partie, on relie la notion de compressibilité à la notion de cardinalité. La troisième partie traitera du résultat indécidabilité, on montrera qu'un certain problème d'apprentissage EMX n'est pas décidable. Enfin il faudrait idéalement une dernière partie pour traiter de la question de la dimension, mais par manque de temps cette partie ne sera pas traiter, mais l'idée de la preuve est incluse dans la conclusion.

Tout au long de ce document -sauf mention contraire- X désignera un ensemble quelconque, appelé ensemble domaine. Et $\mathcal{F}$ désignera une collection quelconque de parties de X.

1 Relating Compression and Learnability

L'objet de cette partie est d'expliciter la démonstration du premier lemme de l'article ; a savoir l'équivalence entre la Compressibilité faible et l'apprenabilité faible. Il nous faut donc définir les notions de compressibilité et d'apprenabilité. La première sous partie expose la notion de compressibilité d'une famille d'ensemble $\mathcal{F}$.

1.1 Monotone Compression Schemes, Boosting, équivalence compression forte et faible

1.1.1 Definition et Objectif

On parlera de compressibilité relativement à une famille d'ensemble $\mathcal{F}$. L'idée c'est que $\mathcal{F}$ est une classe de sous ensemble du domaine X, les éléments de $\mathcal{F}$ sont dans un contexte d'apprentissage nommés concepts ou hypothèses, dans ce même contexte, toute partie d'une ensemble h de la famille $\mathcal{F}$ sera nommé échantillon, ils seront généralement noté S (pour sample) et on spécifiera le plus souvent leurs tailles.

Exemple 1.1. On pourrais avoir comme domaine $X = \mathbb{R}^2$, et chercher a apprendre la classe des rectangles $\mathcal{F} = \{$ rectangle de $\mathbb{R}^2\}$. Un échantillon serait alors un ensemble de points de $\mathbb{R}^2$ contenu dans un rectangle, donc en faites un ensemble points quelconque de $\mathbb{R}^2$.

Ou bien on pourrait avoir $X = \mathbb{R}$, en cherchant à apprendre la classe des triplets de points $\mathcal{F} = \{\{x_1, x_2, x_3\} \mid x_1, x_2, x_3 \in \mathbb{R}$ et distincts deux à deux $\}$

La notion de compression est la suivante, premièrement, il s'agirait d'avoir une méthode, un schème, qui, a partir d'un ensemble de points finis et ordonné non nécessairement distincts (qu'on pourra représenté par un vecteur), permet de retourner un ensemble de la famille $\mathcal{F}$ notons que la conditions sur les ensembles retourner est assez faible, on attends juste que ceux ci soit dans $\mathcal{F}$ c'est à dire qu'on retourne une hypothèse.

Mais on parlera seulement de compression si cette méthode vérifie de plus que, quelque soit un ensemble h de $\mathcal{F}$, pour toute partie ou échantillon (d'une taille donné) de h, il existe une liste (un vecteur) composée de points de l'échantillon, et normalement d'une taille plus petite que la taille de l'échantillon (c'est le fait que cette taille soit plus petite qui justifie le terme de compression), telle que, en appliquant la méthode sur cette liste, on obtient un nouvel ensemble -hypothèse- qui contient l'échantillon initiale.

Plus formellement, on définit ;

Definition 1.1 (Monotone compression Schemes)**.** On appelle Scheme de compression monotone $m \to d$ de $\mathcal{F}$, toute application $\eta : X^d \to \mathcal{F}$, telle que :

Pour tout element h de $\mathcal{F}$, quel que soit S une partie de m points de h, Il existe v un vecteur de d points de S, tel que $S \subseteq \eta(v)$.

Remarquons à quel point la notion formelle ci dessus peut s'éloigner de l'intuition associer au mot compression. Qu'en est t-il du cas pathologique ou, la taille des entrées d est plus grande que la taille des échantillons compressible k ?

En fait avec l'axiome du choix, ce cas est toujours vrai ; dans un tel contexte, un scheme de compression monotone existe, par exemple il suffirait de prendre l'application $v \mapsto Im(v)$ sur X, et pour chaque échantillon S de taille k, choisir un vecteur qui dans ses coordonnés contient tout les éléments de l'échantillon S.

On aurait alors clairement l'inclusion, et même l'identité de S et de l'ensemble des coordonnées du vecteur, ie. $S = Im(v)$.

Puisque $Im(v) \subseteq S \subseteq h$ l'ensemble $C_v = \{g \in \mathcal{F} \mid Im(v) \subseteq g\}$ est donc non vide, par l'axiome du choix sur la collection d'ensemble $\{C_v\}_{v \in X^d}$ on pourrait alors choisir un tel concept g via une fonction de choix ch. Et l'application $v \mapsto ch(C_v)$ serait alors un bon scheme.

Ce cas sera peu étudier ici, le plus souvent on supposera $d < k$.

Remark (Sur les approches ensembliste et vectorielle)*.* Aussi remarquons une équivalence dans la définition que nous avons donné, un compresseur est une application de X^d vers $\mathcal{F}$, X^d est l'ensemble des vecteurs à d points de X. Il convient de remarquer que la notion de schème de compression monotone est à peu de chose près équivalente, que l'on considère que η de X^d vers $\mathcal{F}$ ou bien de $X^{\leq d}$ (Les parties de X à moins de d éléments) vers $\mathcal{F}$. Explicitons les correspondances canoniques.

Si $\eta : X^d \to \mathcal{F}$ est un schème de compression. Sans rentré dans le détails, par l'axiome du choix on peux a toute partie A de moins de d points associer un vecteur $i(A)$ de X^d. Alors posons η' qui à une partie A de X de moins de d points associe $\eta(i(A))$. Il est clair que η' partage la propriété de compression de η.

A l'inverse si $\eta : \mathcal{P}_{\leq d}(X) \to \mathcal{F}$ possède la propriété de compression. Alors considérons $\eta' : X^d \to \mathcal{F}$ qui à un vecteur v associe l'image par η de l'ensemble des coordonnées de v ; a savoir $\eta(Im(v))$. Dès lors η' partage la propriété de compression que η.

En fonction du contexte on adoptera l'approche ensembliste au celle vectorielle, celles ci étant équivalentes.

Une condition va régulièrement être exigée sur la famille $\mathcal{F}$, c'est le fait que la famille soit dirigée. On la définit comme suit ;

Definition 1.2 (Famille dirigée). On dira qu'une famille $\mathcal{F}$ de partie de X est dirigée, si et seulement si,

Pour tout $h_1, h_2 \in \mathcal{F}$, il existe $h \in \mathcal{F}$ tel que ; $h_1 \cup h_2 \subseteq h$

Et, le premier résultat qu'on va chercher à montrer est le théorème de boosting, il requiert que $\mathcal{F}$ soit dirigée et il s'énonce comme suit ;

Theorem 1.1 (Boosting). *Si la famille $\mathcal{F}$ ne contient que des ensembles finis, et est de plus dirigée. étant donnée $d < m$ deux entiers.*

On a équivalence entre les deux assertions ;

1. *Il existe un $m \to d$ schème de compression monotone de $\mathcal{F}$*
2. *Quel que soit un entier $M \geq m$, Il existe un $M \to d$ scheme de compression monotone de $\mathcal{F}$*

Remarquons que le sens de l'équivalence (2) $\Rightarrow$ (1), est immédiat, et est en fait peu intéressant.

Il s'agira donc de montrer (1) $\Rightarrow$ (2), celle ci exprime le fait que étant donné un schème, une méthode, de compression qui réduit à une taille d les échantillons de taille m, on pourra en fait réduire a la taille d tous les échantillons de taille $M \geq m$.

Précisons cependant que, étant donné $M > m$, la méthode (c'est à dire formellement le schème) qui compressera les échantillons de tailles M à la taille m, n'est pas le même que la même que la méthode qui agit sur les échantillons $m + 1$ et les compresse a la taille m.

La preuve de l'implication consistera donc à partir d'un certain schème, d'en déduire l'existence d'un autre. En fait on va même décrire un procédé, qui à partir d'un schème en construire un second.

Techniquement (et on l'explicitera), l'implication (1) $\Rightarrow$ (2) du théorème de Boosting, se déduit et est même équivalente au lemme suivant. De ce lemme on déduirais le théorème de boosting par récurrence.

Lemma 1.2 (Boosting faible). *Si la famille $\mathcal{F}$ ne contient que des ensembles finis, et est de plus dirigée. étant donnée $d < m$ deux entiers.*

Si, il existe un schème de compression monotone $m \to d$ de $\mathcal{F}$,

alors, il existe également un schème de compression monotone $m + 1 \to d$ de $\mathcal{F}$

L'intérêt de ce lemme, c'est qu'on n'a plus a chercher à compresser des échantillons de tailles $M > m$ quelconque, on va seulement chercher à compresser les échantillons de tailles $m + 1$.

Tous l'enjeu de la suite de cette partie c'est de décrire comment a partir d'une méthode qui compresse les échantillons de taille m, on en construit une qui compresse les échantillons de taille $m+1$.

L'idée de la preuve est la suivante en posant η un schème de compression monotone $m \to d$ de $\mathcal{F}$, et étant donné S une partie de $m+1$ points d'un concept h. On cherche à expliciter à la fois une méthode λ et un sous échantillon S' de S tel que, $S \subseteq \lambda(S')$.

Pour cela, on considère x un points quelconque de S, puis on remarque que $S \setminus \{x\}$ est un échantillon de m points. On peut donc compresser $S \setminus \{x\}$ par η en un sous échantillon T, c'est à dire que, $S \setminus \{x\} \subseteq \eta(T)$.

T étant de taille $d < m$, alors puisque $m-d > 0$, considérons T' une partie de $S \setminus T$ contenant x, et de taille $m-d$. Alors l'union $T \cup T'$ est un sous échantillon de S de taille m, que l'on compresse en un sous échantillon R.

Alors on prétends que R est une compression de S pour une application $\lambda : P \mapsto \lambda(P) = H \in \mathcal{F}$ contenant $\bigcup\limits_{Q \in \eta(P)^{\leq d}} \eta(Q)$.

En effet ;

- $T \subseteq T \cup T' \subseteq \eta(R)$, et T est de taille d d'où $\eta(T) \subseteq \lambda(R)$, mais T est la compression de $S \setminus \{x\}$, $S \setminus \{x\} \subseteq \lambda(R)$
- étant une compression, R est un sous échantillon de d points de $T \cup T' \subseteq \eta(R)$. donc, il en suit que $\eta(R) \subseteq \lambda(R)$, donc à fortiori $T' \subseteq \lambda(R)$, mais T' contient x, d'où $x \in \lambda(R)$

Donc, puisque $\lambda(R)$ contient x et $S \setminus \{x\}$. On conclut donc, $S \subseteq \lambda(R)$.

en fait il faut aussi s'assurer qu'une telle fonction λ existe, c'est l'objet de la partie qui suit, on cherche à montrer comment assurer l'existence des fonction $T \mapsto H_T \in \mathcal{F}$ contenant A_T. Pour cela on va faire appel à l'axiome du choix.

1.1.2 Une condition suffisante de compressibilité avec le choix

Puisque l'on va chercher à utiliser l'axiome du choix, on l'énonce ici ;

Axiome 1 (Axiome du choix)**.** *Etant donné une famille $\mathcal{D}$ d'ensemble non vide, Il existe une fonction $f : \mathcal{D} \to \bigcup \mathcal{D}$ dites de choix telle que, pour tout $A \in \mathcal{D}, f(A) \in A$.*

Ainsi, la proposition suivante découle assez simplement de l'axiome du choix.

Proposition 1.1 (Condition suffisante de sélection)**.** *Étant donné E et F des ensembles quelconques. Soit $\Gamma : E \to \mathcal{P}(F)$ une application qui a un points de E associe une partie de F.*

Si pour tout x dans E, $\Gamma(x)$ est non vide.

Alors, il existe $\gamma : E \to F$ telle que, pour tout $x \in E$, $\gamma(x) \in \Gamma(x)$.

On dira que γ est un sélecteur de Γ.

Démonstration. On va en fait chercher a utiliser l'axiome du choix.

On va considérer donc la famille $\mathcal{D} = \{\Gamma(x) \mid x \in E\}$, c'est une famille d'ensemble. Et en vue de la prémisse, tout ensemble de cette famille est non vide.

$\mathcal{D}$ est une famille d'ensemble non vide, et on peux donc utiliser l'axiome du choix sur la collection $\mathcal{D}$. On en déduit, qu'il existe une fonction de choix ch sur $\mathcal{D}$. C'est a dire une fonction $ch : \mathcal{D} \to \bigcup \mathcal{D}$ telle que, pour tout $A \in \mathcal{D}, ch(A) \in A$.

Considérons maintenant, la composée des fonction suivante :

$$\gamma : \begin{array}{ccccc} E & \longrightarrow & \mathcal{D} & \longrightarrow & \bigcup \mathcal{D} \\ x & \xrightarrow{\Gamma} & \Gamma(x) & \xrightarrow{ch} & ch(\Gamma(x)) \end{array}$$

Montrons que γ est la fonction recherchée. en effet, en considérant $x \in E$, on à $\gamma(x) = ch(\Gamma(x)$. Et par definition de la fonciton de choix, $ch(\Gamma(x) \in \Gamma(x)$, ainsi en identifiant γ on obtient, $\gamma(x) \in \Gamma(x)$.

Ce qui signifie que γ est un sélecteur de Γ. □

Notation 1.1. Étant donné A une partie de X on notera $A_{\mathcal{F}}$ l'ensemble des éléments de $\mathcal{F}$ qui contienne A, autrement dit, $A_{\mathcal{F}} = \{h \in \mathcal{F} \mid A \subseteq h\}$.

Étant donné I un ensemble quelconque et une application $f : I \to \mathcal{P}(X)$, on notera $f_{\mathcal{F}} : I \to \mathcal{P}(\mathcal{F})$ l'application qui à $i \in I$ associe $f(i)_{\mathcal{F}}$.

Corollary 1.2.1. *Soit $f : I \to \mathcal{P}(X)$ Les assertions sont équivalentes ;*

1. *Pour tout $i \in I$, il existe un concept h tel que, $f(i) \subseteq h$.*
2. *Pour tout $i \in I$, $f_{\mathcal{F}}(i)$ est non vide.*
3. *Il existe un sélecteur de $f_{\mathcal{F}}$.*

Démonstration. (1) ⇔ (2) est direct, de même pour (3) ⇒ (1). Enfin pour (2) ⇒ (3)On applique la proposition 1.1. □

Proposition 1.2. *Soit $f : X^d \to \mathcal{P}(X)$ une application, et soit $\overline{f}$ un sélecteur de $f_{\mathcal{F}}$.*

Étant donné m un entier, Si, pour tout $h \in \mathcal{F}$, pour toute partie S de m points de h, il existe v tel que, $S \subseteq f(v)$

Alors $\overline{f}$ est un $m \to d$ schème de compression monotone de $\mathcal{F}$.

Démonstration. On à bien $\overline{f} : X^d \to \mathcal{F}$. Et puisque, $f(v) \subseteq \overline{f}(v)$ pour tout $v \in X^d$. On en déduit que puisque f $(m \to d)$-compresse $\mathcal{F}$ il en suit que $\overline{f}$ $(m \to d)$-compresse aussi $\mathcal{F}$.

On conclut donc que $\overline{f}$ est bien un $m \to d$ schème de compression monotone de $\mathcal{F}$. □

Theorem 1.3 (Condition suffisante de compressibilité.)**.** *Soit $f : X^d \to \mathcal{P}(X)$ une application.*

Étant donné m un entier, si on à les prémisses ;

- *(L'information est toujours recouverte par $\mathcal{F}$) Pour tout $x \in X^d$, il existe un concept h qui recouvre $f(x)$, $f(x) \subseteq h$.*
- *(f compresse les échantillons de taille m) Pour tout $h \in \mathcal{F}$, pour toute partie S de m points de h, il existe v tel que, $S \subseteq f(v)$.*

Alors $\mathcal{F}$ est $m \to d$ compressible par un sélecteur de $f_{\mathcal{F}}$.

Démonstration. On applique le corollaire et la proposition précédente. $\square$

Il convient de remarquer que la proposition et le théorème précédent sont aussi valide pour une fonction quelconque $f : \mathcal{P}_{\leq d}(X) \to \mathcal{P}(X)$

1.1.3 Theoreme de Boosting, Decompresseur

En vue d'utiliser ce théorème posons la définition suivante, on choisit l'approche ensembliste ;

Definition 1.3 (Décompresseur carré)**.** soit $\eta : \mathcal{P}_{\leq d}(X) \to \mathcal{F}$ une application.

On appelle décompresseur carré de η (ou si il n'y a pas de confusion seulement carré), l'application suivante,

$$\eta^{*2} : \mathcal{P}_{\leq d}(X) \to \mathcal{P}(X), \quad T \mapsto \bigcup_{R \in \eta(T)^{\leq d}} \eta(R).$$

Intuitivement, étant donné $\eta : \mathcal{P}_{\leq d}(X) \to \mathcal{F}$ une application, et T une partie de au plus d points de X ; à quoi corresponds un élément de $\eta^{*2}(T)$? L'image par η d'un sous échantillon R, peut s'interpréter, comme la donné $h \in \mathcal{F}$ qui est retourner par η pour les données R.

Et, $\eta^{*2}(T) = \bigcup_{R \in \eta(T)^{\leq d}} \eta(R)$, c'est donc la collection de toute les décompressions par η, depuis le concept $\eta(T)$. Ainsi on décompresse d'abord T en $\eta(T)$ puis on regarde l'information qu'on peut atteindre avec η à partir du concept $\eta(T)$.

On pourrais donc dire qu'on applique deux fois η, pas au sens d'une composée de fonctions mais au sens ou on applique une fois η pour obtenir $\eta(T)$, puis on va contenir toute l'information qu'on peut atteindre depuis $\eta(T)$ en appliquant -comme une seconde fois- η.

On cherche maintenant des conditions suffisante pour que la classe des concepts recouvre l'image de η^{*2}. Explicitons quelque contre exemple, pour justifier les prémisses du théorème suivant.

Exemple 1.2. Dans le cas ou $X = \mathbb{R}$, et $\mathcal{F}$ est la classe des triplets de points réel.

Remarquons que $\mathcal{F}$ n'est pas dirigée. Donc étant donné $T_x = \{x_1, x_2, x_3\}$ et $T_y = \{y_1, y_2, y_3\}$ deux triplets alors, on à pas l'assurance que $T_x \cup T_y$ est un réduit a 3 points, ie. est un concept de $\mathcal{F}$, le seul cas ou cet union est de cardinal 3 c'est si $T_x = T_y$, sinon elle est de cardinal plus que 3 et ne peux donc être contenu dans aucun triplet de points.

Mais on peut remarquer qu'il ne peut pas y avoir de compression $3 \to 2$ de $\mathcal{F}$. Sinon admettons que $\{x_1, x_2, x_4\}$ soit compresser en $\{x_1, x_2\}$, que $\{x_1, x_3, x_5\}$ soit compresser en $\{x_1, x_3\}$, et que $\{x_2, x_3, x_6\}$ soit compresser en $\{x_2, x_3\}$. Alors on ne pourra pas parvenir à compresser $\{x_1, x_2, x_3\}$.

Ceci montre que l'hypothèse qu'il existe un $m + 1 \to m$ schème de compression monotone n'est pas une hypothèse gratuite qui serait toujours vrai.

Suffit t-il d'exiger de $\mathcal{F}$ qu'elle soit une famille dirigée, il s'avère que ce n'est pas une condition suffisante. Prenons le cas ou $X = \mathbb{R}^2$, et $\mathcal{F}$ est la classe des rectangles. Dans ce cas, $\mathcal{F}$ est dirigé, on peux toujours contenir deux rectangle dans un autre. Mais, les Rectangles ne sont pas des ensembles fini de points de

$\mathbb{R}^2$. Alors si on considère une application $\eta : (\mathbb{R}^2)^d \to \mathcal{F}$. Étant donné $v \in (\mathbb{R}^2)^d$, alors $\eta^{*2}(v) = \bigcup\limits_{w \in \eta(v)^d} \eta(w)$.

On ne peut alors pas assurer que $\bigcup\limits_{w \in \eta(v)^d} \eta(w)$ est contenu dans un rectangle, ceci malgré le fait que $\mathcal{F}$ est dirigé. C'est parce-que l'ensemble $\eta(v)$ est un rectangle donc un ensemble infini de points, donc l'ensemble $\{\eta(w) \mid w \in \eta(v)^d\}$ est potentiellement un ensemble infini de rectangle, et on ne peut pas assurer même en s'appuyant sur un argument basé sur le fait que $\mathcal{F}$ dirigée, que la réunion de ces éléments est contenu dans un rectangle.

En effet on pourrait construire un contre exemple, remarquons que tous rectangle est inclus dans un demi plan du type $\mathcal{D}_a = \{(x, y) \in \mathbb{R}^2; y < a\}$, en supposant que $\bigcup\limits_{w \in \eta(v)^d} \eta(w)$ est inclus dans un element de $\mathcal{F}$, ie. un rectangle, on va chercher a mettre en échec la propriété de borne. $\eta(v)^d$ étant infini considérons un sous ensemble dénombrable $\{v_n\}_{n \in \mathbb{N}}$ de vecteur de $\eta(v)^d$. Maintenant faisont une hypothèse sur η, on impose à η que, pour tout entier n, $\eta(v_n)$ est dans le complémentaire du demi plan $\mathcal{D}_n$, c'est à dire dans $\mathcal{D}_n^c$, ceci signifie que $\eta(v_n)$ à une position d'ordonnée supèrieur à n. Alors on se demande si un rectangle peut contenir tous les $\eta(v_n)$, mais c'est impossible car une partie contenant tous ces rectangles est alors non bornée, et une partie qui n'est pas bornée n'est pas un rectangle.

La condition sur $\mathcal{F}$ comme famille dirigée n'est donc pas suffisante. Suffit t-il d'exiger que $\mathcal{F}$ est une famille de partie finie ? Considérons $X = \mathbb{N}$ et la famille $\mathcal{F} = \mathcal{P}_{fin}(\mathbb{N}^*) \cup \{0\}$. Alors l'union $\{0\} \cup \{1\} = \{0, 1\}$ n'est pas un concept de $\mathcal{F}$, et n'est pas inclus dans un concept de $\mathcal{F}$, donc la condition $\mathcal{F}$ est une famille de partie finie n'est pas suffisante, c'est aussi quelque chose qu'on pouvais construire dans le cas des triplets de points.

Theorem 1.4 (Condition suffisante pour que le codomaine soit recouvert par la classe des concepts)**.** *Si $\mathcal{F}$ est une famille dirigée, d'ensemble fini.*
Alors, quel que soit $\eta : \mathcal{P}_{\leq d}(X) \to \mathcal{F}$ une application de $\mathcal{P}_{\leq d}(X)$ vers $\mathcal{F}$;

Pour toute partie d'au plus d points T, $\Delta_\eta(T)$ est contenu dans un élément de $\mathcal{F}$.

Démonstration. Soit T une partie de d points (au plus) de X, on à $\eta^{*2}(T) = \bigcup\limits_{R \in \eta(T)^{\leq d}} \eta(R)$.

Puisque $\eta(T)$ est un élément de $\mathcal{F}$, c'est une partie de cardinal fini, il en suit également que $\eta(T)^{\leq d}$ est fini. De fait , $\{\eta(R) \mid R \in \eta(T)^{\leq d}\}$ est aussi fini, et tous ces éléments sont des concepts de $\mathcal{F}$. De fait puisque $\mathcal{F}$ est dirigé, par induction on en déduit qu'il existe $h \in \mathcal{F}$ contenant $\bigcup\{\eta(R) \mid R \in \eta(T)^{\leq d}\} = \bigcup\limits_{R \in \eta(T)^{\leq d}} \eta(R)$.

Autrement dit, h contient $\eta^{*2}(T)$. $\square$

L'intêret de la notion de décompresseur est par la suite explicite puisqu'elle va nous permettre de renforcer la propriété de compression et donc de monter le théoreme de boosting faible.

Lemma 1.5 (Renforcement de la compression)**.** *Soit $m > d$ deux entiers, étant donnée $\eta : \mathcal{P}_{\leq d}(X) \to \mathcal{P}(X)$ une application.*

Si, η $(m \to d)$-compresse $\mathcal{F}$, Alors η^{*2} $(m+1 \to d)$-compresse $\mathcal{F}$.

Démonstration. Considérons donc h un concept, et S un ensemble de $m+1$ points de h.

Soit x un points quelconque de S, dès lors, $S \setminus \{x\}$ est de taille m, donc on peut le compresser en un sous échantillon T par η, ie. $S \setminus \{x\} \subseteq \eta(T)$.

T est de taille $d < m$. Et $S \setminus T$ est non vide de cardinal $m+1-d > 0$, alors il existe T' une partie de $S \setminus T$ contenant x de cardinal $m-d$. alors l'union $T \cup T'$ est de taille m et donc est compressible par η en un sous échantillon R.

Montrons alors que $\eta^{*2}(R)$ contient S.

- $T \subseteq T \cup T' \subseteq \eta(R)$, et T est de taille au plus d donc, $\eta(T)$ est contenu dans $\eta^{*2}(R)$. Alors puisque T est compression de $S \setminus \{x\}$, $S \setminus \{x\} \subseteq \eta^{*2}(R)$.
- Par ailleurs $R \subseteq T \cup T' \subseteq \eta(R)$, et R est de taille d donc, $\eta(R) \subseteq \eta^{*2}(R)$, et alors puisque R est une compression de $T \cup T'$ et que $x \in T'$, on à $x \in \eta^{*2}(R)$.

Donc on conclut que $S \subseteq \eta^{*2}(R)$.

□

Alors en appliquant le théorème 1.3 on va pourvoir déduire le théorème de boosting.

Lemma 1.6 (Boosting faible)**.** *Si la famille $\mathcal{F}$ ne contient que des ensembles finis, et est de plus dirigée. étant donnée $d < m$ deux entiers.*

Si, il existe un schème de compression monotone $m \to d$ de $\mathcal{F}$,

alors, il existe également un schème de compression monotone $m+1 \to d$ de $\mathcal{F}$

Démonstration. Si on considère donc $\eta : \mathcal{P}_{\leq d}(X) \to \mathcal{F}$ un $m \to d$ scheme de compression de $\mathcal{F}$.

Le codomaine de η est $\mathcal{F}$ et $\mathcal{F}$ est par hypothèse une famille dirigé d'ensemble finie, donc en appliquant le théorème 1.4, on en déduit que, toute image de η^{*2} est contenu dans un concept de $\mathcal{F}$.

Par ailleurs, η $(m \to d)$-compresse $\mathcal{F}$, donc par le lemme précédent on en déduit que, η^{*2} $(m+1 \to d)$-compresse $\mathcal{F}$.

De ces deux assertions on en déduit par le théorème 1.3, que $\mathcal{F}$ est $m+1 \to d$ compressible. □

Enfin on peut déduire aisément le théorème de boosting.

Theorem 1.1 (Boosting)**.** *Si la famille $\mathcal{F}$ ne contient que des ensembles finis, et est de plus dirigée. étant donnée $d < m$ deux entiers.*

On a équivalence entre les deux assertions ;

1. *Il existe un $m \to d$ schème de compression monotone de $\mathcal{F}$*
2. *Quel que soit un entier $M \geq m$, Il existe un $M \to d$ scheme de compression monotone de $\mathcal{F}$*

Démonstration. On a évidemment (2) $\Rightarrow$ (1) puisque $m \geq m$.

Pour montrer (1) $\Rightarrow$ (2) il suffit de raisonner par récurrence sur k pour montrer qu'il existe un scheme de compression $m + k \to d$ pour $\mathcal{F}$, pour tout entier naturel k.

Et pour cela on applique simplement le lemme de Boosting faible ; car d'une compression $m + k \to d$ on peut alors en déduire l'existence d'une compression $m + k + 1 \to d$. □

1.1.4 Compression faible et compression forte

Definition 1.4 (Compression Faible)**.** On dit que la famille $\mathcal{F}$ est faiblement compressible, si et seulement si,

Il existe un entier m tel que, $\mathcal{F}$ admet un scheme de compression monotone $m + 1 \to m$.

Definition 1.5 (Compression Forte)**.** On dit que la famille $\mathcal{F}$ est fortement compressible, si et seulement si,

Il existe un entier d, tel que pour tout entier m, il existe un $m \to d$ monotone compression scheme de $\mathcal{F}$.

Alors par le boosting on peut remarquer que ces notions sont équivalentes pour une famille $\mathcal{F}$ d'ensemble finie et dirigée, on à ;

Theorem 1.7. *Si $\mathcal{F}$ est une famille d'ensemble finie et dirigée, les assertions suivantes sont équivalentes ;*

1. *$\mathcal{F}$ est faiblement compressible.*
2. *$\mathcal{F}$ est fortement compressible.*

1.2 EMX Learner

1.2.1 Definition of Learner

Étant donné P une loi de probabilité sur X on notera, $Opt_P(\mathcal{F})$ pour signifier $\underset{h\in\mathcal{F}}{sup}\mathbb{E}_P(h)$.

On définit ici la notion d'EMX learner, EMX signifiant ici "Estimating The Maximum".

Notons que dans cette partie l'approche vectorielle des schèmes peut être nécessaire, puisque que l'intêret est qu'une suite de variable aléatoires est aussi une variable aléatoire, tandis qu'un ensemble de partie aléatoire est une notion sur laquelle on ne peut pas dire autant de chose en probabilité.

Definition 1.6 (EMX-Learner)**.** On appelle (ε, δ)-EMX Learner de la famille $\mathcal{F}$, toute fonction $G : \bigcup_{m\in\mathbb{N}} X^m \to \mathcal{F}$ telle que :

Il existe un entier d tel que, quel que soit P une loi de probabilité a support fini sur X, pour toute famille i.i.d. $(S_1, ..., S_d)$ de variable aleatoire suivant la loi P on a ;

$$\Pr[\underset{P}{\mathbb{E}}(G(S_1, ..., S_d) \leq \underset{P}{Opt}(\mathcal{F}) - \varepsilon] \leq \delta \quad (\star)$$

Maintenant si on écrit,

$$error_{G,P,\mathcal{F}}(\varepsilon) = \{x \in X^d \mid \underset{P}{Opt}(\mathcal{F}) - \underset{P}{\mathbb{E}}(G(x)) \geq \varepsilon\}$$

alors $(\star)$ s'écrit ;

$$S * \mathbb{P}[error_{G,P,\mathcal{F}}(\varepsilon)] \leq \delta$$

Quelques définitions sur la terminologie d'apprenabilité forte et faible,

Definition 1.7 (Apprenabilité faible et apprenabilité)**.** On dira que $\mathcal{F}$ est (ε, δ)-EMX apprenable, si il existe un (ε, δ)-EMX Learner de $\mathcal{F}$.

En particulier dans l'article on parle d'apprenabilité faible si $\mathcal{F}$ est $(\frac{1}{3}, \frac{1}{3})$-EMX apprenable.

On dira que $\mathcal{F}$ est EMX apprenable (on parlera parfois d'apprenabilité forte), si il existe $G : \bigcup_{m\in\mathbb{N}} X^m \to \mathcal{F}$ tel que pour tout ε, δ, G est un (ε, δ)-EMX Learner de $\mathcal{F}$.

Étant donné $f \in \mathcal{F}$ pour faire sens de la quantité $\mathbb{E}_P(f) = \int_X f dP$, il faut que f soit mesurable pour la mesure P. On fait donc les hypothèses suivantes ;

Hypothesis (Hypothèse de mesurabilité)**.** *Dans ce document, toute les lois de probabilité P sont des mesures de l'espace mesurable $(X, \mathcal{P}(X))$ et sont à support fini.*

Alors puisque toute les parties de X sont mesurables, en particulier tout les concepts $f \in \mathcal{F}$ sont mesurables.

1.2.2 Extension du domaine d'une application

Ici, la demarche est la même on va chercher à utiliser le theorème 1.3 pour ca, on va faire appelle a la fonction suivante, ce n'est pas tout a fait la même que précédemment.

Definition 1.8 (Extension d'une application)**.** Étant donné une fonction $f : \mathcal{P}_{\leq d}(X) \to \mathcal{P}(X)$ une application, et m un entier. On définit l'extension de f à la taille m comme suit ;

$$f_{(m)} : \begin{array}{ccc} \mathcal{P}_{\leq m}(X) & \to & \mathcal{P}(X) \\ T & \to & \begin{cases} \text{si T est contenu dans un concept h} ; \; T \cup \bigcup\limits_{R \in T^{\leq d}} f(R) \\ \text{sinon} ; \; \bigcup\limits_{R \in T^{\leq d}} f(R) \end{cases} \end{array}$$

Ici la condition de recouvrement est moins forte que pour le premier type de fonctions. Il suffit que $\mathcal{F}$ soit dirigée.

Proposition 1.3 (Condition de recouvrement)**.** *Si $\mathcal{F}$ est une famille dirigée alors, quel que soit $f : \mathcal{P}_{\leq d}(X) \to \mathcal{F}$.*

Pour tout entier m, le codomaine de $f_{(m)}$ est recouvert par $\mathcal{F}$.

Autrement dit, pour toute partie T à moins de m points de X, $f_{(m)}(T)$ est contenu dans un élément de $\mathcal{F}$.

Démonstration. Soit T une partie de m points de X, on va considérer deux cas.

1. Si T n'est pas contenu dans un élément de $h \in \mathcal{F}$ alors par définition, $f_{(m)}(T) = \bigcup\limits_{R \in T^{\leq d}} f(w)$.

Puisque T est finie, donc $T^{\leq d}$ [2] est finie, et donc il en suit que $\{f(R) \mid R \in T^{\leq d}\}$ est de même finie.

Alors puisque $\mathcal{F}$ est dirigée on en déduit que il existe un concept $g \in \mathcal{F}$ contenant $\bigcup\{f(R) \mid R \in T^{\leq d}\} = \bigcup\limits_{R \subseteq T, |R| \leq d} f(R) = f_{(m)}(T)$.

2. Maintenant dans le cas où T est contenu dans un élément $h \in \mathcal{F}$, alors $f_{(m)}(T) = T \cup \bigcup\limits_{R \subseteq T, |R| \leq d} f(R) \subseteq h \cup g$. Et puisque $\mathcal{F}$ est dirigée, on en déduit que $h \cup g$ est contenu dans un élément $h' \in \mathcal{F}$, donc on en déduit à fortiori,

$$f_{(m)}(T) \subseteq h'.$$

□

Maintenant on se demande comment une extension se comporte relativement à la propriété de compression.

Lemma 1.8 (Conséquence de la non compressibilité)**.** *Considérons d un entier quelconque $G : \mathcal{P}_{\leq d}(X) \to \mathcal{F}$ une application.*

Quel que soit $d < m$ un entier, si l'extension $G_{(m)}$ ne $(m+1 \to m)$-compresse pas la famille $\mathcal{F}$,

Alors Il existe $h \in \mathcal{F}$, et il existe S un ensemble de $m+1$ points de h, tels que quel que soit T le sous échantillon de d point de S, on à ;

2. On note $T^{\leq d}$ pour $\mathcal{P}_{\leq d}(T)$ l'ensemble des parties finies de T à au plus d éléments.

1. *Pour tout $x \in S$, si T ne contient pas d'occurrence de x, alors x n'appartient pas a $G(T)$, ie, $x \notin G(T)$*
2. *$G(T) \cap S \subseteq T$*
3. *pour tout vecteur $v \in X^d$, $\mathbb{E}_P(G_{X^d}(v)) \leq \frac{d}{m+1}$ où P est la distribution uniforme sur S, et $G_{X^d}(v) = G(Im(v))$*

Démonstration. Supposons donc que $G_{(m)}$ ne $(m+1 \to m)$-compresse pas la famille $\mathcal{F}$.

C'est à dire que il existe un $h \in \mathcal{F}$, et S un ensemble de $m+1$ points de h, tels que, quel que soient S' un sous échantillon de m points de S, S n'est pas contenu dans $G_{(m)}(S')$.

Alors, soit x un élément quelconque de S. Considérons en particulier le sous échantillon $S' = S \diagdown \{x\}$ de S. Il en suit, par l'hypothèse, que en particulier, S n'est pas inclus dans $G_{(m)}(S')$, ie. $S \nsubseteq G_{(m)}(S')$.

Alors, remarquons que, S' est un sous échantillon de m points de $S \subseteq h$, de fait il existe donc un concept h contenant S'. Et alors, $G_{(m)}(S')$ contient S', c'est à dire $S \diagdown \{x\}$.

Pour mettre en échec $S \subseteq G_{(m)}(S')$, il faut donc que nécessairement, $x \notin G_{(m)}(S')$, donc à fortiori $x \notin \bigcup_{T \subseteq S', |T| \leq d} G(T)$.

C'est à dire que x n'appartient à aucune image $G(T)$, ou T est un sous échantillon de d points de $S' = S \diagdown \{x\}$.

1. C'est a dire donc que pour tout sous échantillon T de d points de S dans lequel il n'y a pas d'occurrence de x ; $x \notin G(T)$. Donc on a montrer que, pour tout sous échantillon T de d points de S, pour tout $x \in S$ si x n'apparait pas dans T alors, $x \notin G(T)$

2. Soit $x \in S$ en considérant T un sous échantillon de d de S, si $x \notin T$ alors c'est équivalent au fait que T est un échantillon de d points de $S \setminus \{x\}$, ce qui par ce le point précédent entraîne, $x \notin G(T)$.

Donc on montrer l'implication $x \notin T \cap S \Rightarrow x \notin G(T) \cap S$. Donc par la contraposée on à l'implication $x \in G(T) \cap S \Rightarrow x \in T \cap S = T$. Autrement dit on à l'inclusion $G(T) \cap S \subseteq T$. Et de fait on à aussi l'inégalité des cardinaux $|G(T) \cap S| \leq |T| \leq d$.

3. Alors soit v un vecteur de d points de S, alors remarquons que $Im(v)$ est un sous échantillon de au plus de d points de S. On peux donc appliquer les propositions précédentes, et en déduite $|G(Im(v)) \cap S| \leq d$.

Alors, en faisant appel a la proposition, en posant P la distribution uniforme sur S, on à $\mathbb{E}_P(G_{X^d}(v)) = \frac{|G_{X^d}(v) \cap S|}{|S|}$ [3], et donc avec l'inégalité précédente, $\frac{|G(Im(v)) \cap S|}{|S|} \leq \frac{d}{m+1}$, d'où $\mathbb{E}_P(G_{X^d}(v)) \leq \frac{d}{m+1}$. □

3. Lorsque $f : \mathcal{P}_{\leq d}(X) \to \mathcal{P}(X)$ on note f_{X^d} pour la correspondance canonique entre approche vectorielle et ensembliste, $f_{X^d}(v) = f(Im(v))$.

1.2.3 L'apprenabilité faible entraîne la compressibilité faible

Remark (Correspondance entre l'approche vectorielle et ensembliste). Étant donné $G : \bigcup_{m\in\mathbb{N}} X^m \to \mathcal{F}$ une application, comment faire la correspondance de l'approche vectorielle a celle ensembliste. Remarquons que on peut restreindre G à l'espace X^d. Ensuite par l'axiome du choix on peut établir une bijection $i : \mathcal{P}_{\leq d}(X) \to X^d$. Alors, $G_{|X^d} \circ i : \mathcal{P}_{\leq d}(X) \to \mathcal{F}$, on pourra donc sur cette fonction appliquer les propositions précédentes.

Argument pour l'existence de la bijection i. il suffit de remarquer que $X^d \to \mathcal{P}_{\leq d}(X), v \to Im(v)$ injecte X^d dans $\mathcal{P}_{\leq d}(X)$. Que par ailleurs étant donné, A une partie de au plus d points de X, on peut indexer A et lui associer un vecteur de la famille $\overline{A} = \{u \in X^d \mid A = Im(u)\}$. Alors on peux injecter $\mathcal{P}_{\leq d}(X)$ dans X^d en appliquant le choix sur la collection $\{\overline{A} \mid A \in \mathcal{P}_{\leq d}(X)\}$. On pourrait alors par exemple conclure l'existence de la bijection par Cantor Bernstein.

Corollary 1.8.1 (Decompresseur d'un appreneur et compressibilité)**.** *Considérons $G : \bigcup_{m\in\mathbb{N}} X^m \to \mathcal{F}$ un $(\frac{1}{3}, \frac{1}{3})$-EMX learner de $\mathcal{F}$, de paramètre $d = d(\frac{1}{3}, \frac{1}{3})$. En considérant $m = \lfloor \frac{3d}{2} \rfloor \geq d$.*

Nécessairement alors, l'application $[G_{|X^d} \circ i]_{(m)}$ $(m+1 \to m)$-compresse la famille $\mathcal{F}$.

Démonstration. On va raisonner par l'absurde. Supposons que $[G_{|X^d} \circ i]_{(m)}$ ne $m+1 \to m$ compresse pas $\mathcal{F}$, alors on à exactement les prémisses de la propositions précédentes. On en déduit donc l'existence d'un élément $h \in \mathcal{F}$ et d'une partie S de $m+1$ points de h, tels que, quel que soit v un vecteur de d points de S, on à, $\mathbb{E}_P([G_{|X^d} \circ i]_{X^d}(v)) \leq \frac{d}{m+1}$. Où P la distribution uniforme sur S.

Mais, $[G_{|X^d} \circ i]_{X^d}(v) = G_{|X^d}(i[Im(v)]) = G(i[Im(v)])$. Donc l'assertion affirme que ; quel que soit v un vecteur de d points de S, on à,

$$\mathbb{E}_P[G(i[Im(v)])] \leq \frac{d}{m+1}.$$

Mais puisque i est une bijection de X^d sur $\mathcal{P}_{\leq d}(X)$, et que toute partie de $\mathcal{P}_{\leq d}(X)$ peut s'écrire comme l'image d'un vecteur de d points. Autrement dit, quelque soit u un vecteur de d points de X, u peut s'écrire comme une image de i, $u = i(A)$ où A est une partie d'au plus d points de X, mais A peut aussi s'écrire comme l'image d'un vecteur $v \in X^d$, ie.$A = Im(v)$, et alors $u = i(Im(v))$.

On en déduit alors qu'en fait, quel que soit v un vecteur de d points de S ;

$$\mathbb{E}_P(G(v)) \leq \frac{d}{m+1}$$

Remarquons que Avec P la distribution uniforme sur S, on à $Opt_P(\mathcal{F}) = 1$.

Considérons alors $(Ti)_{1\leq i\leq d}$ une famille de d variables aléatoires, indépendantes et identiquement distribuées, suivant la loi P. On veut mesurer $\{T \in error_{G,P,\mathcal{F}}(\frac{1}{3})\}$ [4], pour cela montrons que $\{T \in S^d\}$ est inclus dans $\{T \in error_{G,P,\mathcal{F}}(\frac{1}{3})\}$.

Considérons donc ω un element de $\{T \in S^d\}$. C'est donc que en particulier, le vecteur $T(\omega) = (T_1(\omega), ..., T_d(\omega))$ est un vecteur de d points de S. Par le point

4. Etant donné A une partie de X, on note $\{T \in A\}$ pour l'image réciproque $\{\omega \in \Omega \mid T(\omega) \in A\}$

précédent, on en déduit donc que, $\mathbb{E}_P[G(T(\omega))] \leq \frac{d}{m+1}$.

Et c'est la où la valeur de m joue son rôle, $\frac{d}{m+1} = \frac{d}{\lfloor\frac{3d}{2}\rfloor+1} \leq \frac{d}{\frac{3d}{2}} = \frac{2}{3}$. Alors il en suit que $\mathbb{E}_P(G(T(\omega))) \leq \frac{2}{3}$ on en déduit donc,

$$\mathbb{E}_P(G(T(\omega))) - Opt_P(\mathcal{F}) = \mathbb{E}_P(G(T(\omega))) - 1 \leq \frac{2}{3} - 1 = -\frac{1}{3}.$$

Et donc de facon équivalente, $Opt_P(\mathcal{F}) - \mathbb{E}_P(G(T(\omega))) \geq \frac{1}{3}$, c'est donc par definition que, $T(\omega)$ est un element de $error_{G,P,\mathcal{F}}(\varepsilon)$. On viens donc de montrer que

$$\{T \in S^d\} \subseteq \{T \in error_{G,P,\mathcal{F}}(\tfrac{1}{3})\}.$$

Il en suit donc l'inégalité des mesures,

$$\mathbb{P}(\{T \in S^d\}) \leq \mathbb{P}(\{T \in error_{G,P,\mathcal{F}}(\tfrac{1}{3})\}).$$

Mais, $\mathbb{P}(\{T \in S^d\}) = 1$, et donc $\mathbb{P}$ étant une mesure de probabilité on en déduit finalement que $\mathbb{P}(\{T \in error_{G,P,\mathcal{F}}(\frac{1}{3})\}) = 1$.

De fait alors,

$$\mathbb{P}(\{T \in error_{G,P,\mathcal{F}}(\tfrac{1}{3})\}) = \mathbb{P}_T(error_{G,P,\mathcal{F}}(\tfrac{1}{3})) = 1 > \tfrac{1}{3}.$$

Ceci contredit donc le fait que G est un$(\frac{1}{3}, \frac{1}{3})$-EMX learner de $\mathcal{F}$, de paramètre $d = d(\frac{1}{3}, \frac{1}{3})$.

Par l'absurde, on conclut donc que nécessairement $[G_{|X^d} \circ i]_{(m)}$ $(m+1 \to m)$-compresse la famille des concepts $\mathcal{F}$ □

Theorem 1.9 (L'Apprenabilité faible implique la compressiblité faible)**.** *Si la famille $\mathcal{F}$ est dirigée. Quel que soit $G : \bigcup_{m\in\mathbb{N}} X^m \to \mathcal{F}$ un $(\frac{1}{3}, \frac{1}{3})$-EMX learner de $\mathcal{F}$, de paramètre $d = d(\frac{1}{3}, \frac{1}{3})$. En considérant $m = \lfloor\frac{3d}{2}\rfloor \geq d$.*

Nécessairement alors, la famille $\mathcal{F}$ est $m+1 \to m$ compressible.

Démonstration. On cherche à utiliser le théorème 1.3.

- Premièrement on applique la proposition 1.3, pour assurer que le codomaine de $[G_{|X^d} \circ i]_{(m)}$ est recouvert par la classe $\mathcal{F}$ des concepts.
- Par ailleurs, on applique le corollaire 1.8.1, qui permet d'assurer que $[G_{|X^d} \circ i]_{(m)}$ $m+1 \to m$ compresse $\mathcal{F}$.

On en déduit alors que $\mathcal{F}$ est $m+1 \to m$ compressible (par un scheme de compression monotone). □

1.3 How compression implies learnability

L'objet de cette partie est de montrer le theoreme suivant ;

Theorem 1.10 (La compression faible entraîne l'apprenabilité EMX forte)**.** *Si $\mathcal{F}$ est une collection d'ensemble fini, et est dirigée, étant donnée m un entier quelquonque.*

Il existe un entier $M \geq m$ tel que, si la famille $\mathcal{F}$ est $M \to m$ compressible, alors $\mathcal{F}$ est EMX apprenable. C'est à dire qu'il existe $G : \bigcup_{m\in\mathbb{N}} X^m \to \mathcal{F}$ tel que pour tout ε, δ, G est un (ε, δ)-EMX Learner de $\mathcal{F}$.

Autrement dit, la compressibilité faible entraîne l'apprenabilité EMX forte.

Démonstration. admise, cf. partie 8 de l'article [2], on peux voir aussi le theoreme 2.1 de [6] ou bien le théorème 1.2 de [7], qu'il faut "adapter" au cas de l'apprentissage EMX. □

Remarquons alors ce qui suit de ce théorème, on a montrer que la compressibilité forte et faible sont équivalente. De plus on à montrer que l'apprenabilité faible entraîne la compressibilité faible. Alors en fait ce théorème montrera le suivant ;

Theorem 1.11 (Caractérisation générale)**.** *Si $\mathcal{F}$ est une collection d'ensemble fini, et est une famille dirigée. Les assertions suivantes sont équivalentes ;*

1. *$\mathcal{F}$ est faiblement compressible.*
2. *$\mathcal{F}$ est fortement compressible.*
3. *$\mathcal{F}$ est faiblement EMX apprenable.*
4. *$\mathcal{F}$ est fortement EMX apprenable.*

Pour l'équivalence (3) $\Leftrightarrow$ (4), on parle aussi de boosting (comme c'était le cas pour la compression). On dit qu'on boost un appreneur faible en un appreneur fort.

2 Relating Compression and Cardinality

L'objet de cette partie est lier les notions de cardinalité de la famille $\mathcal{F}$ et la notion de compressibilité de $\mathcal{F}$. Cependant dans cette partie $\mathcal{F}$ n'est plus quelconque, $\mathcal{F}$ corresponds à l'ensemble des parties finies de X, on notera cet ensemble $X^{<\mathbb{N}}$, $X^{<\omega}$ ou plus simplement $\mathcal{P}_{fin}(X)$. On notera ω est le cardinal de l'ensemble $\mathbb{N}$ cf. à l'annexe B pour la définition formelle de ω.

Plus précisement dans cette partie on veut montrer le résultat suivant ;

Theorem 2.1. *Etant donné un entier k, et X un ensemble. Les assertions suivantes sont équivalentes ;*

1. *La cardinalité de X est au plus, $\aleph_k$, ie. $card(X) \leq \aleph_k$*
2. *Il existe un scheme de compression monotone $(k+2) \to (k+1)$ de la famille $\mathcal{P}_{fin}(X)$*

On montrera l'équivalence en deux temps, comme deux implications (1) $\Rightarrow$ (2) et (2) $\Rightarrow$ (1). Lors de cette partie on va utiliser formellement la notion de cardinal, on va donc souvent se référer à l'annexe B. On va commencer par montrer l'implication dans le sens indirect (2) $\Rightarrow$ (1).

2.1 La compressibilité entraîne une majoration de la cardinalité

2.1.1 Existence de points irrécouvrables

Definition 2.1 (Point irrécouvrable)**.** Soit k un entier positif non nul, et X un ensemble de cardinal infini. Étant donné η un $k+1 \to k$ schème de compression monotone de $\mathcal{P}_{fin}(X)$,

On dit qu'un point x de X est irrécouvrable par η sur une partie Y de X, si et seulement si,

Quel que soit v un vecteur de k points de Y, $\eta(v)$ ne contient pas x, ie. $x \notin \eta(v)$

De façon équivalente, cela signifie, $x \notin \bigcup_{v \in Y^k} \eta(v)$

Lemma 2.2 (Cardinalité de l'union des images par un schème)**.** *Soit k un entier positif non nul, et X un ensemble de cardinal infini.*

Si, il existe η un $k+1 \to k$ scheme de compression monotone de $\mathcal{P}_{fin}(X)$,

Quel que soit Y une partie de X de cardinal infini ; Y a le même cardinal que $\bigcup_{v \in Y^k} \eta(v)$

Démonstration. Remarquons que quel que soit v un vecteur de k points de Y ; $\eta(v)$ est une partie finie de X. C'est à dire par le corollaire B.2.1 que $|\eta(v)| < \aleph_0$.

Donc $\aleph_0$ majore l'ensemble $\{card(\eta(v)) : v \in Y^k\}$, il est donc plus petit que, la borne supérieur (ie. le plus petit majorant).

C'est à dire, $sup\{card(\eta(v)) : v \in Y^k\} \leq \aleph_0$.

Mais, Y est de cardinal infini, donc par le même corollaire, $\aleph_0 \leq card(Y)$, on en déduit donc que, $sup\{card(\eta(v)) : v \in Y^k\} \leq card(Y)$.

De plus, puisque Y est de cardinal infini, par le corollaire $B.2.3$ il en suit que Y^k et Y ont le même cardinal. C'est à dire, $card(Y^k) = card(Y)$.

Et donc $sup\{card(\eta(v)) : v \in Y^k\} \leq card(Y^k)$.
D'où, $sup[card(Y^k), sup\{card(\eta(v)) : v \in Y^k\}] = card(Y^k) = card(Y)$

Alors remarquons que par la proposition $B.18$ on à,
$\left|\bigcup_{v\in Y^k} \eta(v)\right| \leq card(Y^k \times sup\{card(\eta(v)) : v \in Y^k\})$.

Mais, Y^k étant un ensemble infini, par la proposition $B.17$ le cardinal de ce produit cartésien est, $sup[card(Y^k), sup\{card(\eta(v)) : v \in Y^k\}]$
Mais, on à vu que ce cardinal est en fait $card(Y)$

D'où : $\left|\bigcup_{v\in Y^k} \eta(v)\right| \leq |Y| \qquad (i)$

Montrons que Y est inclus dans $\bigcup_{v\in Y^k} \eta(v)$. Soit y un élément de Y. Y est de cardinal infini on peut construire, ou plutôt, il existe, une partie finie S de Y composée de $k+1$ points et contenant y.
Et par la nature de η, il existe un vecteur u de k points de S, tel que, $S \subseteq \eta(u)$. Et donc de fait, $y \in \eta(u)$, ou u est un vecteur de k points de S donc de Y, c'est donc que $u \in Y^k$. D'où enfin, $y \in \bigcup_{v\in Y^k} \eta(v)$.

Ainsi on a montrer que $Y \subseteq \bigcup_{v\in Y^k} \eta(v)$. Donc il en suit l'inégalité des cardinaux, $|Y| \leq \left|\bigcup_{v\in Y^k} \eta(v)\right| \qquad (ii)$

On conclut par (i) et (ii) ; $\quad |Y| = \left|\bigcup_{v\in Y^k} \eta(v)\right|$ □

Corollary 2.2.1 (Existence de point irrécouvrable)**.** *Soit k un entier positif non nul, et X un ensemble de cardinal infini. Si, il existe η un $k+1 \to k$ schème de compression monotone de $\mathcal{P}_{fin}(X)$,*

Quel que soit Y une partie de X de cardinal infini, si $|Y| < |X|$;

Alors il existe $x \in X$ un point irrécouvrable par η sur Y, c'est à dire tel que $x \notin \bigcup_{v\in Y^k} \eta(v)$.

Démonstration. On à exactement les même prémisses que le lemme précédent, on peut donc l'appliquer et déduire que ; Y a le même cardinal que $\bigcup_{v\in Y^k} \eta(v)$.

mais pourtant $|Y| < |X|$, donc, $\left|\bigcup_{v\in Y^k} \eta(v)\right| < |X|$.

donc nécessairement $\bigcup\limits_{v\in Y^k} \eta(v) \subsetneq X$. C'est à dire qu'il existe, $x \in X$, tel que $x \notin \bigcup\limits_{v\in Y^k} \eta(v)$. □

Remark. Remarquons que ici l'approche vectorielle et ensembliste sont équivalentes puisque pour un ensemble Y quelconque, Y^k et $\mathcal{P}_{\leq k}(Y)$ ont le même cardinal.

2.1.2 Construction de scheme à compression plus forte, sur une partie stricte

Ici on utilise l'approche ensembliste des schèmes de compression monotone.

Lemma 2.3. *Étant donné X un ensemble, η un $k+1 \to k$ scheme de compression monotone de $\mathcal{P}_{fin}(X)$, , Y une partie de X, et $x \in X$ un point irrécouvrable par η sur Y.*

Pour toute partie T de k éléments de Y ; quel que soit T' un sous échantillon de k points de $T \cup \{x\}$ tel que, $T \cup \{x\} \subseteq \eta(T')$
x apparait dans le sous échantillon T', $x \in T'$.

Démonstration. Montrons que x est dans T'.
En effet si $x \notin T'$ puisque, $T' \subseteq T \cup \{x\}$, on en deduit, que T' est un sous échantillon de k points de T, mais T est une partie de Y. donc, T' est un échantillon de k points de Y. C'est à dire $T' \in Y^{\leq k}$.

Mais par hypotheses sur x, $x \notin \bigcup\limits_{R\in Y^{\leq k}} \eta(R)$.
A fortiori on en déduirais donc que, $x \notin \eta(T')$.
Mais c'est absurde puisque, $T \cup \{x\} \subseteq \eta(T')$ donc à fortiori, $x \in \eta(T')$.
Par l'absurde on conclut ; $x \in T'$. □

Theorem 2.4. *Soit k un entier positif non nul, et X un ensemble de cardinal infini. Y une partie de X, de cardinal infini :*

Quel que soit η un $k+1 \to k$ scheme de compression monotone de $\mathcal{P}_{fin}(X)$, et $x \in X$ un point irrécouvrable par η sur Y.
Alors, l'application suivante est un scheme de compression monotone $k \to k-1$ de $Y^{\leq \mathbb{N}}$,

$$\lambda : \begin{array}{ccc} Y^{k-1} & \longrightarrow & Y^{<\mathbb{N}} \\ R & \longrightarrow & \eta[R \cup \{x\}] \cap Y \end{array}$$

Démonstration. Montrons que λ est un scheme pour la famille $\mathcal{P}_{fin}(Y)$. Considérons donc h une partie finie de Y et T une partie finie de k points de h (donc de Y).

Remarquons que, $T \cup \{x\}$ est une partie de $k+1$ points de X, et, η est un schème de X. Donc il existe T' un sous-échantillon de k points de $T \cup \{x\}$ tel que, $T \cup \{x\} \subseteq \eta(T')$.
De par le lemme précédent on en déduit que, x apparait dans le vecteur T', c'est à dire que $x \in T'$. il en suit que, $x \in \eta(T')$ et donc $T \cup \{x\} \subseteq \eta(T')$ de fait

$$T \cup \{x\} \cap Y = T \subseteq \eta(T') \cap Y$$

Enfin considérons $R = T' \setminus \{x\}$. Alors R est un sous échantillon de taille $k-1$ de T et l'inclusion précédente s'écrit alors ;

$$T \subseteq \eta(R \cup \{x\}) \cap Y$$

autrement dit,

$$T \subseteq \lambda(R)$$

où R est un sous échantillon de T de taille $k-1$. On à donc compresser T par λ un échantillon quelconque de k points de Y en un sous échantillon de taille $k-1$.

□

Corollary 2.4.1. *Soit k un entier positif non nul, et X un ensemble de cardinal infini, et η un $k+1 \to k$ scheme de compression monotone de $\mathcal{P}_{fin}(X)$,*

Quel que soit Y une partie de X, de cardinal infini telle que $card(Y) < card(X)$,

Il existe un schème de compression monotone $k \to k-1$ de $Y^{<\mathbb{N}}$,

Démonstration. Il suffit d'appliquer le corollaire 2.2.1 pour obtenir un point irrécouvrable, puis appliquer le théorème précédent. □

2.1.3 Un ensemble de cardinal trop grand ne peut pas être compresser

Proposition 2.1 (Schème de compression $1 \to 0$ d'un ensemble infini)**.** *Quel que soit X un ensemble,*

Si X est de cardinal infini, il n'existe pas de schème de compression monotone $1 \to 0$ de $\mathcal{P}_{fin}(X)$

Démonstration. On va raisonner par l'absurde ; Soit η un schème de compression monotone $1 \to 0$ de $X^{\leq \mathbb{N}}$. η est une application de $\mathcal{P}_{\leq 0}(X)$ dans $X^{\leq \mathbb{N}}$. $\mathcal{P}_{\leq 0}(X)$ n'a qu'un élément, l'ensemble vide $\emptyset$. Donc η n'a qu'une image ; $\eta(\emptyset)$.

Soit $x \in X$, $\{x\}$ est une partie de $\mathcal{P}_{fin}(X)$ à un élément, elle est donc compressible par η. C'est à dire qu'il existe un sous échantillon T de taille 0 de $\{x\}$ et tel que, $\{x\} \subseteq \eta(T)$. Or le seul sous échantillon de taille nulle c'est l'ensemble vide $\emptyset$. D'où, $\{x\} \subseteq \eta(\emptyset)$, ie. $x \in \eta(\emptyset)$.

D'où l'inclusion $X \subseteq \eta(\emptyset)$. Mais par définition de η est sensé vérifier, $\eta(\emptyset)$ dans $\mathcal{P}_{fin}(X)$, c'est à dire que $\eta(\emptyset)$ est un ensemble fini. Mais X est infini et inclus dans $\eta(\emptyset)$. D'où la contradiction. □

Corollary 2.4.2. *Soit k un entier positif, et X un ensemble.*

Si X est de cardinal infini $\aleph_{k+1}$, alors il n'existe pas de schème de compression monotone $k+2 \to k+1$ de $\mathcal{P}_{fin}(X)$,

Démonstration. Par récurrence sur k.

Si $k = 0$: Par l'absurde, soit η un scheme de compression monotone $2 \to 1$ de $\mathcal{P}_{fin}(X)$. X est de cardinal $\aleph_1$, donc Il existe Y une partie de X de cardinal $\aleph_0$. Et alors, $card(Y) < card(X)$ et Y est de cardinalité infini.

On peut donc appliquer le corollaire 2.4.1 sur Y et en déduire un schème de compression monotone $1 \to 0$ de $\mathcal{P}_{fin}(Y)$.

Mais Y est infini puisque de cardinal $\aleph_0$. Donc ceci viens contredire la proposition 2.1.

Supposons l'assertion vrai au rang k[5] et montrons la pour $k+1$. Par l'absurde, supposons qu'il existe η un schème de compression monotone $k+3 \to k+2$ de $\mathcal{P}_{fin}(X)$ avec X de cardinal $\aleph_{k+2}$, alors il existe Y une partie de X de cardinal $\aleph_{k+1}$.

On peux appliquer le corollaire 2.4.1 sur Y et en déduire un schème de compression monotone $k+2 \to k+1$ de $\mathcal{P}_{fin}(Y)$.

Mais ceci viens alors contredire l'hypothèse de récurrence au rang k. □

Alors remarquons que étant donné $Y \subseteq X$ deux ensembles, si $\mathcal{P}_{fin}(X)$ est $k+2 \to k+1$ compressible alors il en suit que $\mathcal{P}_{fin}(Y)$ est de même $k+2 \to k+1$ compressible.

Alors avec le dernier résultat on en déduit l'implication suivante ;

$$card(X) \geq \aleph_{k+1} \Rightarrow X \text{ n'est pas } k+2 \to k+1 \text{ compressible.}$$

Donc la contraposée nous permet d'avoir une des implication du théorème qu'on souhaite montrer ;

Corollary 2.4.3. *Soit k un entier positif, et X un ensemble.*
Si, $\mathcal{P}_{fin}(X)$ est $k+2 \to k+1$ compressible,
Alors, X est de cardinal $\aleph_k$ au plus, ie. $card(X) \leq \aleph_k$

5. C'est à dire qu'on suppose que, quel que soit E un ensemble de cardinal $\aleph_{k+1}$ il n'existe pas de scheme de compression monotone $k+2 \to k+1$ de $\mathcal{P}_{fin}(E)$.

2.2 Un ensemble de cardinalité suffisamment petite, est compressible

On veut montrer l'implication dans le sens direct du theoreme :

Theorem 2.1. *Etant donné un entier k, et X un ensemble. Les assertions suivantes sont équivalentes ;*

1. *La cardinalité de X est au plus, $\aleph_k$, ie. $card(X) \leq \aleph_k$*
2. *Il existe un scheme de compression monotone $(k+2) \to (k+1)$ de la famille $\mathcal{P}_{fin}(X)$*

autrement dit on va vouloir montrer la proposition suivante ;

Theorem 2.5. *Etant donné k un entier naturel, quelque soit X un ensemble de cardinalité au plus $\aleph_k$; $\mathcal{P}_{fin}(X)$ est $k+2 \to k+1$ compressible.*

Dans un premier temps il convient de remarquer une propriété de la compressibilité de la famille $\mathcal{P}_{fin}(X)$, en fait en compressant un échantillon S en un sous échantillon S'. Lorsque l'on décompresse S' l'information de S' est de fait deja accessible puisque S' est une partie finie de X, il faut donc surtout s'attarder à trouver comment retrouver l'information de $S \setminus S'$.

On formule ici plus formellement ce qu'on entends par là ;

Proposition 2.2 (Caractérisation de la compressibilité de la famille des parties finies)**.** *Étant donné X un ensemble*

1. *Il existe un scheme de compression $k \to d$ monotone η de $\mathcal{P}_{fin}(X)$.*
2. *Il existe $f : \mathcal{P}_{\leq k}(X) \to \mathcal{P}_{fin}(X)$ tel que, Pour toute partie S de k points de X, il existe S' un sous échantillon de d points de S tel que, $S \setminus S' \subseteq f(S')$*

Démonstration. Il est clair que (1) $\Rightarrow$ (2) puisque étant donné S un échantillon de taille k et S' un sous échantillon de S, on à $S \setminus S' \subseteq S \subseteq \eta(S')$. Donc η à la propriété recherchée.

Pour (2) $\Rightarrow$ (1), il suffit de poser $\eta : T \mapsto f(T) \cup T$, η est alors un $k \to d$ scheme de compression monotone de $\mathcal{P}_{fin}(X)$. □

En fait la preuve rédigée dans l'article consiste exactement à construire une fonction qui vérifie la propriété 2 de la caractérisation, c'est pour ça qu'on l'explicite ici, et aussi car la rédaction de la preuve par induction est plus aisée avec cette caractérisation.

2.2.1 Cas initial, cas d'un ensemble au plus dénombrable.

Pour montrer le théorème 2.5, on va raisonner par récurrence sur l'entier k. Traitons d'abord le cas ou k est nulle.

On va admettre le théorème suivant, qui par ailleurs est équivalent à l'axiome du choix ;

Theorem 2.6. *(Zermelo) Tout ensemble peut être muni d'un bon ordre*

et on va notamment utiliser le corollaire de ce théorème

Corollary 2.6.1. *Tout ensemble E peut être muni d'un bon ordre $<$, tel que, $(E, <)$ est de type d'ordre un cardinal.*

C'est à dire que en notant α le cardinal de E, $(E, <)$ et $(\alpha, \in)$ sont isomorphes au sens du bon ordre.

Notons que ce corollaire signifie que le fait que pour un ensemble quelconque E supposé l'existence d'un bon ordre $<$ sur E tel que, $(E, <)$ est de type d'ordre un cardinal, est une hypothèse gratuite, dans le sens ou c'est toujours vrai. Dans un énoncé on pourra donc toujours ajouté dans les prémisses, un tel ordre $<$ pour tout ensemble E, et cette prémisse ne "coûte" rien, si on parvient a démontrer un theoreme avec cette premisse, on l'a aussi démontrer sans cette premisse.

Proposition 2.3 (Cas dénombrable)**.** *Soit X un ensemble de cardinalité au plus $\aleph_0$.*

Étant donné $<$ un bon ordre sur X, tel que $(X, <)$ est de type d'ordre un cardinal.

Alors $f : X \to \mathcal{P}_{fin}(X)$ qui à x associe S_x [6], vérifie que quel que soit S une partie de 2 points de X, il existe S' un sous échantillon de 1 points de S tel que, $S \setminus S' \subseteq f(S')$

Et donc par la proposition 2.2 ceci équivaut au fait que $\mathcal{P}_{fin}(X)$ est $2 \to 1$ compressible.

Démonstration. **1. Bonne définition.** Montrons que f est bien définit de X vers $\mathcal{P}_{fin}(X)$, soit x un element de X. Alors, $f(x) = S_x = \{a \in X \mid a < x\}$ alors nécessairement $x \notin S_x$ (par antisymétrie de $<$). Donc S_x est un segment initial propre de $(X, <)$.

Par ailleurs $(X, <)$ est de type d'ordre un cardinal, qui est $\omega = \omega_0$.

Donc en appliquant le théorème B.4, on en déduit, $card(S_x) < card(X) = \aleph_0$, S_x est donc fini, ie. $S_x \in \mathcal{P}_{fin}(X)$.

Et donc en identifiant, $f(x) \in \mathcal{P}_{fin}(X)$.

2. Soit S un échantillon de 2 points, alors $S = \{x, y\}$. Considérons que x est le max pour $<$ de S (quitte à permuter x et y) alors, $y < x$ donc $y \in S_x$ c'est à dire, $S \setminus \{x\} = \{y\} \subseteq S_x$ autrement, $S \setminus \{x\} \subseteq f(x)$.

C'est bien la propriété recherchée. □

2.2.2 Stabilité par passage au successeur

On veut montrer maintenant la stabilité par passage au successeur, c'est à dire que étant donné k un entier naturel non nul. C'est à dire qu'on veut montrer que ;

Si, quel que soit X un ensemble de cardinal $card(X) \leq \aleph_{k-1}$, il existe un $k+1 \to k$ scheme de compression monotone de $\mathcal{P}_{fin}(X)$.

Alors, pour tout ensemble X de cardinal $card(X) \leq \aleph_k$, il existe un $k+2 \to k+1$ scheme de compression monotone de $\mathcal{P}_{fin}(X)$

Fixons k un entier naturel non nul, et supposons donc dans la partie qui suit que l'assertion suivante est vrai ;

6. S_x c'est le segment initial de l'ensemble bien ordonnée $(X, <)$ suivant ; $\{a \in X \mid a < x\}$

Hypothesis. *Quel que soit X un ensemble de cardinal $card(X) \leq \aleph_{k-1}$, il existe un $k+1 \to k$ scheme de compression monotone de $\mathcal{P}_{fin}(X)$*

Maintenant il s'agit de montrer que, quel que soit X un ensemble de cardinal $card(X) \leq \aleph_k$, il existe un $k+2 \to k+1$ scheme de compression monotone de $\mathcal{P}_{fin}(X)$.

Mais par la proposition 2.2 il suffit de montrer que ; Il existe $f : \mathcal{P}_{\leq k+1}(X) \to \mathcal{P}_{fin}(X)$ telle que quel que soit S une partie de $k+2$ points de X, il existe S' un sous échantillon de $k+1$ points de S tel que, $S \setminus S' \subseteq f(S')$

Naturellement on va chercher à utiliser l'hypothèse de récurrence, il s'agit donc d'expliciter une partie de X de cardinal au plus $\aleph_{k-1}$. On trouve une telle partie en faisant appel au théorème B.4, qui affirme qu'un segment initial propre d'un cardinal α est toujours de cardinalité strictement inférieur à celle de α.

Dans notre cas, remarquons que, pour tout $x \in X$, $S_x = \{y \in x \mid y < x\}$ est un segment initial propre de X, donc par application du théorème B.4, $S_x < card(X) = \aleph_k$, c'est a dire que S_x est de cardinalité au plus $\aleph_{k-1}$.

Alors on peut appliquer l'hypothèse de récurrence et affirmer qu'il existe f_x un $k+1 \to k$ schème de compression monotone de $\mathcal{P}_{fin}(S_x)$.

Enfin par l'axiome du choix, on peux affirmer alors, qu'il existe une application $f : x \to f_x$ qui à x associe un schème de compression monotone de $\mathcal{P}_{fin}(S_x)$.

Remarquons aussi que étant donné que $<$ est un bon ordre sur X, pour toute partie finie S il existe un maximum, qu'on notera $max(S)$ ou $M(S)$

Alors on peux montrer la stabilité par passage au successeur ;

Theorem 2.7 (Stabilité)**.** *Si on à l'hypothèse de récurrence c'est à dire que pour tout ensemble E de cardinal $card(E) \leq \aleph_{k-1}$, il existe un $k+1 \to k$ scheme de compression monotone de $E^{<\omega}$*

Alors, étant donné X un ensemble de cardinal $card(X) \leq \aleph_k$, et étant donné $<$ un bon ordre sur X, tel que $(X, <)$ est de type d'ordre un cardinal.

$g : T \mapsto f_{max(T)}[T \setminus \{max(T)\}]$, $\mathcal{P}_{\leq k+1}(X) \to \mathcal{P}_{fin}(X)$ est un $k+2 \to k+1$ scheme de construction complémentaire de $\mathcal{P}_{fin}(X)$

Démonstration. Soit S une partie de $k+2$ points de X. On va cherche à expliciter un sous échantillon S' de $k+1$ points de S tel que, $S \setminus S' \subseteq g(u)$.

Soit $M(S)$ le maximum pour l'ordre $<$ de S, alors remarquons que par le lemme ; $S_{M(S)}$ admet un scheme de construction complémentaire, $f_{M(S)}$.

De plus, $S \setminus \{M(S)\}$ est une partie du segment initial $S_{M(S)}$, et, cette partie est composée de $k+1$ points, on peut donc appliquer la propriété de compression de $f_{M(S)}$.

Il existe donc S' un sous échantillon de k points de $S \setminus \{M(S)\}$ tel que, $S \setminus \{M(S)\} \subseteq f_{M(S)}(S')$, donc à fortiori $S \setminus \{M(S)\} \setminus S' \subseteq f_{M(S)}(S')$, autrement dit ;

$$S \setminus (\{M(S)\} \cup S') \subseteq f_{M(S)}(S') \quad (\star)$$

Posons $R = \{M(S)\} \cup S'$ et montrons que R est bien une compression de S pour g.

remarquons que, S' ne possède aucune occurrence de $M(S)$ puisqu'il est compose de points de $S \setminus \{M(S)\}$; R est donc un sous échantillon de S de taille $k+1$.

de plus $S' \subseteq S$, donc $max(S) = max(S' \cup \{max(S)\})$, de fait $R \setminus max(R) = S' \cup \{max(S)\} \setminus \{max(S)\} = S'$. Et par ailleurs, $f_{max(S)} = f_{max(R)}$.

Alors on peut écrire $(\star)$ comme ;

$$S \setminus R \subseteq f_{M(R)}(R \setminus \{M(R)\})$$

Et donc,

$$S \setminus R \subseteq g(R).$$

□

Alors sous l'hypothèse de récurrence on à montrer le théorème précédent, mais par la proposition 2.2 on à donc montrer de facon équivalent que, quel que soit X un ensemble de cardinal $card(X) \leq \aleph_k$, il existe un $k+2 \rightarrow k+1$ scheme de compression monotone de $\mathcal{P}_{fin}(X)$.

Donc en fait plus généralement on montrer par induction que, étant donné k un entier naturel ; Quelque soit X un ensemble de cardinalité au plus $\aleph_k$; il existe f un $k+2 \rightarrow k+1$ schème de compression monotone de la famille $\mathcal{P}_{fin}(X)$.

C'est donc qu'on à montrer le théorème 2.5.

2.2.3 Synthèse

On à donc montrer le théorème suivant ;

Theorem 2.5. *Etant donné k un entier naturel, quelque soit X un ensemble de cardinalité au plus $\aleph_k$; $\mathcal{P}_{fin}(X)$ est $k+2 \rightarrow k+1$ compressible.*

Enfin on peut donc conclure l'équivalence entre la compressibilité de la famille des partie finie de X et majoration du cardinal de X.

Theorem 2.1. *Etant donné un entier k, et X un ensemble. Les assertions suivantes sont équivalentes ;*

1. *La cardinalité de X est au plus, $\aleph_k$, ie. $card(X) \leq \aleph_k$*
2. *Il existe un scheme de compression monotone $(k+2) \rightarrow (k+1)$ de la famille $\mathcal{P}_{fin}(X)$*

Démonstration. C'est donc simplement la synthèse du corollaire 2.4.3 et du théorème 2.5. □

3 Undecidability Of The EMX learnability

Maintenant, il s'agit d'utiliser le théorème 2.1 pour montrer l'indécidabilité d'un problème EMX.

Theorem 2.1. *Etant donné un entier k, et X un ensemble. Les assertions suivantes sont équivalentes ;*

1. *La cardinalité de X est au plus, $\aleph_k$, ie. $card(X) \leq \aleph_k$*
2. *Il existe un scheme de compression monotone $(k+2) \to (k+1)$ de la famille $\mathcal{P}_{fin}(X)$*

Alors ce résultat s'applique en particulier pour $X = [0, 1]$, c'est à dire qu'on à,

Proposition 3.1. *Pour tout entier k, les assertions sont équivalentes ;*

1. *$[0, 1]$ est de cardinalité au plus $\aleph_k$.*
2. *Il existe un scheme de compression $k + 2 \to k + 1$ de $\mathcal{P}_{fin}([0, 1])$*

En assumant alors que ZFC (il nous faut donc definir la théorie ZFC) admet deux modèles (c'est aussi donc une notion à définir) suivants ;

(M_0) Dans le premier l'hypothèse du continu est vrai autrement, $2^{\aleph_0} = \aleph_1$

(M_1) Dans l'autre l'hypothèse du continu est fausse et même pour tout ordinal $k < \omega$, $\aleph_k < 2^{\aleph_0}$

On va chercher un montrer un résultat d'indécidabilité.

3.1 Resultats classiques de cardinalité

Theorem 3.1. *Les ensembles $[0, 1]$ et $\mathbb{R}$ partage la même cardinalité.*

Démonstration. **1.** Du fait que $[0, 1]$ est une partie $\mathbb{R}$, on peut naturellement l'injecter dans $\mathbb{R}$.

2. On chercher à injecter $\mathbb{R}$ dans $[0, 1]$, on injecte $\mathbb{R}$ dans $\mathbb{R}^+$ par l'exponentielle $x \mapsto e^x$. Par translation, $x \mapsto x + 2$ on injecte $\mathbb{R}^+$ dans $[2, \infty[$. Puis par la fonction inverse, $x \mapsto 1/x$ on injecte $[2, \infty[$ dans $[0, \frac{1}{2}]$ qu'on peut bien injecter par une identité dans $[0, 1]$.

On conclut par Cantor-Bernstein l'équipotence de $[0, 1]$ et $\mathbb{R}$ $\square$

Le théoreme suivant justifie la notation $2^{\aleph_0}$ pour désigner le cardinal de $\mathbb{R}$, c'est à dire également celui de $[0, 1]$.

Theorem 3.2. *Les ensembles $\mathbb{R}$ et $\mathcal{P}(\mathbb{N})$ ont le même cardinal.*

Démonstration. Préalablement il convient de remarquer que, $\mathcal{P}(\mathbb{N})$ et $\{0, 1\}^{\mathbb{N}}$ sont équipotents. Ceci, par l'application, $A \mapsto \mathbb{1}_A$ qui à une partie associe son indicatrice.

1. On construit une injection de $\{0, 1\}^{\mathbb{N}}$ dans $[0, 1[$, en associant à une suite $\{b_i; i \in \mathbb{N}\}$ le réel $0, b_0 b_1 b_2 ...$ il parait clair que, cette application une injection.

$[0, 1[$ et $\mathbb{R}$ étant équipotents, on en déduit que, $\mathcal{P}(\mathbb{N})$ est subpotent à $\mathbb{R}$.

2. Il s'agit de montrer $\mathbb{R}$ est subpotent à $\mathcal{P}(\mathbb{N})$. Pour se faire, remarquons que $\mathbb{N}$ et $\mathbb{Q}$ sont équipotents, donc il en va de même pour $\mathcal{P}(\mathbb{N})$ et $\mathcal{P}(\mathbb{Q})$. Alors remarquons que l'application $\mathbb{R} \to \mathcal{P}(\mathbb{Q})$, $x \mapsto \{q \in \mathbb{Q} \mid q < x\}$ est injective, ceci est du à la densité de $\mathbb{Q}$ dans $\mathbb{R}$, qu'on peut énoncer comme le fait que entre deux réels, il existe toujours un rationnel.

Alors il en suit que $\mathbb{R}$ est subpotent à $\mathcal{P}(\mathbb{Q})$ donc à $\mathcal{P}(\mathbb{N})$.

On conclut donc (par Cantor-Bernstein) que, $\mathbb{R}$ et $\mathcal{P}(\mathbb{N})$ sont équipotents, c'est à dire qu'ils ont le même cardinal.

$\square$

On cherche à montrer l'indécidabilité de l'apprenabilité EMX de la famille $\mathcal{P}_{fin}([0,1])$.

Remarquons que dans (M_0), $[0,1]$ est de cardinalité $\aleph_1$, donc en appliquant l'equivalence, $\mathcal{P}_{fin}([0,1])$ est $3 \to 2$ compressible, ceci implique alors que, $\mathcal{P}_{fin}([0,1])$ est EMX-apprenable.

Par ailleurs dans (M_1), pour tout entier k, $[0,1]$ est de cardinalité $2^{\aleph_0} > \aleph_k$. Autrement par l'équivalence, pour tout entier non nul k, $\mathcal{P}_{fin}([0,1])$ n'est pas $k+1 \to k$ compressible. On en déduit que, nécessairement $\mathcal{P}_{fin}([0,1])$ n'est pas EMX-apprenable, en utilisant le resultat de la partie 1.

3.2 Formule indépendante, ZFC, et déduction

On veux montrer que l'énoncé suivant $\mathfrak{f}$ est indépendant de ZFC ;

$$\mathfrak{f} = \text{"}\textit{La famille } \mathcal{P}_{fin}([0,1]) \textit{ des parties finies de } [0,1] \textit{ est EMX-apprenable"}$$

Ainsi on aura montrer qu'un problème d'apprentissage EMX est indécidable à partir des axiomes de ZFC. Explicitons ce qu'on entends par indécidabilité, et aussi qu'entends t-on par ZFC.

3.2.1 Formule indépendante, Axiomes de ZFC

Definition 3.1 (Formule indépendante, formule indécidable)**.** On dit qu'une formule ϕ est indépendante d'une théorie T, si et seulement si, T ne prouve pas ϕ et T ne prouve pas $\neg\phi$.

Autrement dit, Il existe un modèle $\mathfrak{M}_1$ de T qui ne satisfait pas ϕ ie. $\mathfrak{M}_1 \not\models \phi$. Et il existe un modèle $\mathfrak{M}_2$ de T qui ne satisfait pas $\neg\phi$ ie. $\mathfrak{M}_2 \not\models \neg\phi$.

Et donc de façon équivalente en se référent a la définition C.9 ; Il existe un modèle de T satisfaisant ϕ et il existe un modèle de T satisfaisant $\neg\phi$.

On dit aussi que la formule ϕ est indécidable dans la théorie T.

On note ZFC pour "Zermelo Fraenkel Choix". Les noms de Zermelo et Fraenkel (ZF) renvoie à des axiomes (des énoncés) de la théorie des ensembles, le Choix renvoie a un seul axiome ; l'axiome du choix.

ZF est une collection d'une dizaine d'énoncé du premier ordre, c'est donc au sens de la définition formelle en logique, une théorie (De même pour ZFC). Ce sont des énoncés du premier ordre c'est à dire que ce sont des formules closes du premier ordre, par exemple un des axiomes de ZF est l'axiome d'extensionnalite, qui corresponds à la formule ;

$$\forall x \forall y[x = y \Leftarrow (\forall z(z \in x \Leftrightarrow z \in y))]$$

Tous les axiomes de ZFC s'énonce de la même façon comme des formules du premier ordre, on peux avoir plus de détails sur ceux ci par exemple avec la référence [10].

En ayant introduit la notion de formules dans un langage du premier ordre, les formules modélise les énoncés mathématiques par exemple l'énoncé qu'on à noter $\mathfrak{f}$ peut être représenté par une formule close du premier ordre.

On aimerais maintenant modéliser les preuves du documents, et plus généralement toutes preuves mathématiques. Pour cela en fait, on admet que l'on peut a toute preuve associer une preuve formelle, c'est à dire une preuve dans un système formelle, par exemple en déduction naturelle ou en calcul des séquents. Et alors il convient de remarquer qu'une preuve formelle d'un énoncé caractérise les conséquences d'une théorie, c'est l'objet du théorème suivant.

Theorem 3.3 (Correction et complétude, déduction naturelle, calcul des séquents)**.** *Etant donné T une théorie et ϕ une formule close. Les assertions sont équivalentes ;*

1. *T prouve ϕ, c'est à dire que tous les modèles de T satisfont ϕ.*
2. *Il existe une preuve en calcul des séquents de $\Gamma \vdash \phi$ avec $\Gamma \subseteq T$, et Γ fini.*
3. *Il existe une preuve en déduction naturelle de $\Gamma \vdash \phi$ avec $\Gamma \subseteq T$, et Γ fini.*

Démonstration. Admis. □

On fait plusieurs hypothèses pour assurer que la théorie des modèles et les langages du premier ordre modélise bien le langage mathématique.

Tout les énoncés E (propositions, lemmes, théorèmes, corollaires) de ce document peuvent êtres formulé comme une formule close du premier ordre ϕ_E, on confondra les objets E et ϕ_E.

A toutes les preuves du document on peux associer un arbre de déduction naturelle (ou de façon équivalente du calcul des séquents) de $\Gamma \vdash \phi$, avec toujours des hypothèses $\Gamma \subseteq ZFC$. C'est à dire que lorsqu'on à démontrer une proposition ϕ on de à façon équivalente, montrer $ZFC \vdash \phi$.

3.3 Indécidabilité d'un probleme d'apprentissage EMX

3.3.1 Modèles de ZFC, Les modèles de Gödel et Cohen

On va admettre deux résultat de théories des modèles, car ceux ci seraient trop long et sans doute trop compliquer à démontrer, on se réfère à [13] ;

Theorem 3.4 (Gödel)**.** *Il existe un modèle $\mathfrak{M}_\mathfrak{o}$ de ZFC, tel que,*
$\mathfrak{M}_\mathfrak{o} \vDash (2^{\aleph_0} = \aleph_1)$

Remarquons que l'indécidabilité de l'hypothèse du continu $2^{\aleph_0} = \aleph_1$ ne sera pas suffisante ici. En fait on fera plutôt appel de l'indécidabilité de la formule $2^{\aleph_0} \leq \aleph_\omega$. Puisque, En effet par le théorème **??** et 2.1 on à les équivalences ;

$$\begin{aligned}
\mathcal{P}_{fin}([0,1]) \text{ est EMX apprenable} \quad &\Leftrightarrow \quad \text{il existe un entier k tel que,} \\
&\qquad \mathcal{P}_{fin}([0,1]) \text{ est } k+1 \to k \text{ compressible} \\
&\Leftrightarrow \quad \text{il existe un entier k tel que,} \\
&\qquad [0,1] \text{ est de cardinal au plus } \aleph_{k-1} \\
&\Leftrightarrow \quad \text{il existe un entier k tel que, } 2^{\aleph_0} \leq \aleph_{k-1} \\
&\Leftrightarrow \quad 2^{\aleph_0} \leq \aleph_\omega (\text{Par croissance des alephs})
\end{aligned}$$

Le modèle de Gödel satisfaisant l'hypothèse du continu, donc il satisfait à fortiori l'énoncé $2^{\aleph_0} \leq \aleph_\omega$.

En revanche pour construire un modèle dans lequel l'énoncé $2^{\aleph_0} \leq \aleph_\omega$ est faux, il faut faire appel à d'autres outils. Premièrement le Forcing de Cohen permet de mettre en échec l'hypothèse du continu, mais pour parvenir à mettre en échec l'énoncé $2^{\aleph_0} \leq \aleph_\omega$, il faut itérer ce processus de Forcing, sans rentrer dans le détail il semblerait que le résultat suivant peut par exemple se déduire d'un résultat associer à Solovay et Martin.

Theorem 3.5 (Cohen-Solovay). *Il existe un modèle $\mathfrak{M}_1$ de ZFC, tel que,* $\mathfrak{M}_1 \vDash (2^{\aleph_0} > \aleph_\omega)$

Remarquons les implications dans chacun des modèles, ceci revient plus ou moins à expliciter l'équivalence écrit précédemment ;

Proposition 3.2 (Conséquence de l'hypothèse de continu). *Si l'hypothèse du continu est vrai alors, la famille $\mathcal{P}_{fin}([0,1])$ est EMX-apprenable.*

Démonstration. Supposons que l'hypothèse du continue est vrai, c'est à dire, $2^{\aleph_0} = \aleph_1$

Mais alors, notons que $[0,1]$ est de cardinalité $2^{\aleph_0}$. Par les théorèmes 3.1 et 3.2.

Il en suit donc que $[0,1]$ est de cardinalité $\aleph_1$. Donc par le théorème 2.1 on en déduit que de facon équivalente, ceci signifie que, $[0,1]^\omega$ est $3 \to 2$ compressible.

Mais de l'assertion $[0,1]^\omega$ est $3 \to 2$ compressible on en déduit par le théorème **??**, que $\mathcal{P}_{fin}([0,1])$ est EMX-apprenable. □

Proposition 3.3 (Dans le modèle de Cohen-Solovay). *Si l'assertion $2^{\aleph_0} > \aleph_\omega$ est vraie, alors la famille $\mathcal{P}_{fin}([0,1])$ n'est pas EMX-apprenable.*

Démonstration. Supposons que $2^{\aleph_0} > \aleph_\omega$. Alors en remarquant le théorème **??**, considérons m un entier naturel, par l'absurde en supposant que $[0,1] < \omega$ est $m+2 \to m+1$ compressible, en vertu du théorème 2.1 ceci signifie de facon équivalente que, $[0,1]$ est de cardinal $\aleph_m$.

Mais par le theorème 3.1, $[0,1]$ est de cardinalité $2^{\aleph_0}$ et donc par hypothèse de cardinal strictement supérieur $\aleph_\omega$.

Par ailleurs, au sens des ordinaux on à $m < \omega$ et donc nécessairement, $\aleph_m < \aleph_\omega$.

On à donc une contradiction $card[0,1] < card[0,1]$. Donc nécessairement la famille des parties finies $[0,1]^{<\omega}$ n'est pas $m+2 \to m+1$ compressible.

De même, On peut exclure le cas $1 \to 0$ par la proposition 2.1. Ainsi on à montrer que quel que soit l'entier m, $[0,1]^{<\omega}$ n'est pas $m+1 \to m$ compressible. Donc par le Théorème **??** on en déduit que la famille, $[0,1]^{<\omega}$ n'est pas EMX-apprenable. □

Sous couvert des remarques sur les langages du première ordre et ZFC, par ces deux propositions on à montrer d'une part $ZFC \vdash (2^{\aleph_0} = \aleph_1) \to \mathfrak{f}$ et d'autres part, $ZFC \vdash (2^{\aleph_0} > \aleph_\omega) \to \neg\mathfrak{f}$

Remarquons alors la proposition suivante ;

Proposition 3.4. *Etant donné T une théorie, ϕ et ψ deux énoncés. Si, T à pour conséquence logique $\phi \to \psi$, et si $T \cup \{\phi\}$ est consistante.*

Alors, $T \cup \{\psi\}$ est consistante. Et même plus précisément tout modèle $\mathfrak{A}$ de T, tel que $\mathfrak{A} \vDash \phi$ vérifie $\mathfrak{A} \vDash \psi$.

Démonstration. $T \cup \{\phi\}$ est consistante donc il existe $\mathfrak{A}$ un modèle de $T \cup \{\phi\}$. Alors en particulier, $\mathfrak{A} \vDash \phi$.

Par ailleurs $T \vdash \phi \to \psi$, et $\mathfrak{A}$ est à fortiori un modèle de T, donc

$$\begin{aligned} \mathfrak{A} \vDash \phi \to \psi &\Leftrightarrow \mathfrak{A} \vDash \neg\phi \vee \psi \\ &\Leftrightarrow \mathfrak{A} \vDash \neg\phi \quad ou \quad \mathfrak{A} \vDash \psi \\ &\Leftrightarrow \mathfrak{A} \nvDash \phi \quad ou \quad \mathfrak{A} \vDash \psi \end{aligned}$$

Mais puisque $\mathfrak{A} \vDash \phi$ on à nécessairement, $\mathfrak{A} \vDash \psi$. Ainsi, $\mathfrak{A}$ est un modèle de $T \cup \{\phi\}$, cette théorie admet donc un modèle, c'est à dire qu'elle est consistante. □

3.3.2 Le résultat d'indécidabilité

Alors on en déduit le théorème suivant, l'indécidabilité de l'apprenabilité EMX pour la famille $\mathcal{P}_{fin}([0,1])$.

Theorem 3.6 (Résultat d'indépendance)**.** *L'énoncé $\mathfrak{f}$ est indépendant de la théorie ZFC.*

C'est à dire que l'énoncé "La famille $\mathcal{P}_{fin}([0,1])$ des parties finies de $[0,1]$ est EMX-apprenable" n'est pas décidable.

Démonstration. Par les theorèmes de Gödel et de Cohen on à l'existence de deux modèles de ZFC $\mathfrak{M}_0$ et $\mathfrak{M}_1$ tels que ;

1. $\mathfrak{M}_0 \vDash (2^{\aleph_0} = \aleph_1)$
2. $\mathfrak{M}_1 \vDash (2^{\aleph_0} < \aleph_\omega)$

Mais les proposition suivantes ont montrées que ;

1. $ZFC \vdash (2^{\aleph_0} = \aleph_1) \to \mathfrak{f}$
2. $ZFC \vdash (2^{\aleph_0} < \aleph_\omega) \to \neg\mathfrak{f}$

On en déduit donc en appliquant la proposition précédente que ;

$$\mathfrak{M}_0 \vDash \mathfrak{f} \quad et \quad \mathfrak{M}_1 \vDash \neg\mathfrak{f}$$

où $\mathfrak{M}_0$ et $\mathfrak{M}_1$ sont des modèles de ZFC. On à donc montrer que, $\mathfrak{f}$ est indépendante des axiomes de ZFC. □

D'ou l'existence d'un problème qui n'est pas EMX apprenable. Ce résultat est assez central dans l'article, puisqu'il viens justifier le titre de l'article "Learnability can be undecidable" ou dans l'autre version "On a learning problem that is independent of the set theory ZFC axioms".

Conclusion

Il reste à traiter la question de la dimension pour l'apprentissage EMX, c'est à dire est ce qu'il existe une dimension qui comme la dimension VC caractérise l'apprentissage PAC, caractériserait de la même façon l'apprentissage EMX. Il s'avère que ce n'est pas le cas, ceci est lié au résultat d'indécidabilité de la partie 3. On explicite la raison pour laquelle une telle dimension ne peut exister, en se refèrant à la partie 6 de l'article [2] "On the existence of a combinatorial dimension for EMX learning".

En fait, on exigerait d'une telle dimension la caractérisation suivante ; $\mathcal{F}$ est EMX-apprenable si et seulement si $\mathcal{F}$ est de dimension $\mathcal{D}(\mathcal{F})$ finie. Mais de plus on exigerait que l'assertion "$\mathcal{D}(\mathcal{F}) \geq d$" peut être démontrée par un nombre fini de points de X et un nombre fini d'éléments de $\mathcal{F}$. On dirait que la propriété "$\mathcal{D}(\mathcal{F}) \geq d$" est de caractère fini. Alors en admettant que l'assertion suivante est vraie ;

Assertion. Étant donné A une propriété de caractère fini, $\mathfrak{M}_0$ et $\mathfrak{M}_1$ deux modèles de ZFC, si $\mathfrak{M}_0$ est un sous modèle de $\mathfrak{M}_1$, et si X et $\mathcal{F}$ sont les mêmes dans les deux modèles ; alors A est vrai dans $\mathfrak{M}_0$ si et seulement si A est vrai dans $\mathfrak{M}_1$.

Sous réserve que cette assertion est vraie, en remarquant que le modèle Cohen est une extension de celui de Gödel, et donc par itération des extensions il en suit que le modèle qu'on a nommé modèle de Cohen-Solovay est aussi une extension de celui de Gödel. Alors en appliquant cette assertion il est clair qu'il ne peut exister une telle dimension $\mathcal{D}$. Puisque dans le modèle de Gödel, $\mathcal{P}_{fin}([0,1])$ est apprenable donc $\mathcal{D}(\mathcal{P}_{fin}([0,1]))$ est fini, de fait il existe un entier d tel que l'énoncé $\mathcal{D}(\mathcal{P}_{fin}([0,1])) \geq d$ est faux. Mais par ailleurs dans le modèle de Cohen-Solovay, $\mathcal{P}_{fin}([0,1])$ n'est pas apprenable donc sa dimension n'est pas fini, donc à fortiori l'énoncé $\mathcal{D}(\mathcal{P}_{fin}([0,1])) \geq d$ est vrai. Alors il est clair qu'en appliquant l'assertion on à une contradiction ; c'est pourquoi il n'existe pas une notion de dimension pour l'apprentissage EMX.

Au vue de cette impossibilité, il est naturelle de s'interroger sur les raisons qui font que les apprentissages EMX et PAC diffèrent sur l'existence d'une notion de dimension. On peut alors remarquer qu'il y a deux différences majeures entre ces deux apprentissages.

La première c'est que lors d'un apprentissage PAC les échantillons sont étiquetés par un concept $h \in \mathcal{F}$, ce ne sont pas simplement des parties finies d'un concept quelconque, ce sont des suites de points étiquetés. Un point étiqueté par un concept $h \in \mathcal{F}$ c'est la donné d'un point et la donné de son appartenance à h. Formellement les échantillons dans l'apprentissage PAC sont des vecteurs $\langle (x_1, h(x_1)), ..., (x_n, h(x_n)) \rangle$, ou h est vue comme son indicatrice. Remarquons que ce n'est pas le cas pour l'apprentissage EMX, il se pourrait que la différence de ces apprentissages soit due à l'étiquetage.

La seconde différence majeure c'est qu'un appreneur EMX n'est pas définit comme une algorithme, mais comme une fonction, ce qui semble générer le problème d'indécidabilité, qui est le résultat centrale de l'article [1].

L'article propose aussi quelques questions ouvertes, notamment sur la partie 1, peut on alléger les hypothèse sur $\mathcal{F}$, peut on montrer que l'apprenabilité entraîne la compressibilité faible sans exiger que $\mathcal{F}$ soit dirigée ? D'après les auteurs, cette question serait lier au problème ouvert de l'apprentissage PAC suivant ; Est ce qu'une classe de concept de dimension VC finie admet toujours un L-W [7] scheme de compression propre, où par *propre* on entendrait que ce scheme puisse reconstruire toute la classe des concepts.

Les auteurs ajoutent une remarque sur le résultat d'indécidabilité pour $\mathcal{P}_{fin}([0,1])$; En fait, l'apprenabilité de cette classe de concept n'est pas un cas réellement susceptible d'être rencontré en pratique dans l'apprentissage automatique. Plutôt ce résultat est là pour montrer que les définitions fondamentales de l'apprentissage PAC ne peuvent pas être généralisées, sans prendre le risque de perdre certaines caractérisations fondamentales.

En conclusion, l'intérêt de la notion de fonction par rapport a celle d'algorithme c'est que l'on à pas à faire intervenir de notion de calcul. Également un des intérêts de cette approche c'est que l'on à pu montrer l'existence d'appreneur, sans expliciter un algorithme, une procédure. En revanche l'article montre que cette approche ensembliste de l'apprentissage possède aussi des inconvénients, notamment l'indécidabilité de l'apprentissabilité EMX de certain problème, mais aussi et par conséquent, on ne peut avoir l'équivalent d'une notion de dimension dans le cadre de l'apprentissage EMX.

7. pour Littlestone-Warmuth, c'est le basic compression scheme de [6].

Annexe A Mesure et Probabilité

A.1 Théorie de la mesure

Definition A.1 (Tribu, Espace Mesurable)**.** On appelle tribu sur un ensemble X, tout ensemble de partie $\mathcal{T}$ de X tel que ;

1. $\mathcal{T}$ contient X
2. $\mathcal{T}$ est stable par passage au complémentaire.
3. $\mathcal{T}$ est stable par union dénombrable.

Alors le couple $(X, \mathcal{T})$ est nommé espace mesurable, on dit parfois en fonction du contexte espace probabilisable.

Aussi on dit que les parties de $\mathcal{T}$ sont des parties $\mathcal{T}$-mesurable ou seulement mesurables de X (si il n'y a pas d'ambiguïté).

Definition A.2 (Mesure, mesure de probabilité)**.** On appelle mesure d'une espace mesurable $(X, \mathcal{T})$, toute application $m : \mathcal{T} \to \mathbb{R}^+$ telle que ;

1. l'ensemble vide est de mesure nulle ; $m(\emptyset) = 0$
2. m est sigma-additive, ie, pour toute famille $(A_i)_{i \in I}$ dénombrable d'éléments de $\mathcal{T}$; $m(\bigcup A_i) \leq \sum m(Ai)$

Alors on dit que $(X, \mathcal{T}, m)$ est un espace mesuré.

on dit qu'une mesure m est une mesure de probabilité si et seulement si, $m(X) = 1$. Dans ce cas on dit que $(X, \mathcal{T}, m)$ est un espace probabilisé.

Definition A.3 (Fonction Mesurable, Variable aléatoire)**.** Soit $(X, \mathcal{A})$ et $(Y, \mathcal{B})$ deux espaces mesurables.

On dit qu'une application $f : X \to Y$ est $(\mathcal{A}, \mathcal{B})$-mesurable, si et seulement si, pour tout $B \in \mathcal{B}$, l'image réciproque de B par f est mesurable, ie. $f^{-1}(B) \in \mathcal{A}$

Alors en particulier si $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé, on dit qu'une application $f : \Omega \to X$ est une variable aléatoire, si et seulement si, f est $(\mathcal{G}, \mathcal{A})$-mesurable.

Dans ce contexte, une image $X(\omega)$ de X est appelée une réalisation de X.

A.2 Loi de probabilité

Proposition A.1 (Loi de probabilité)**.** *Soit* $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (E, \mathcal{E})$ *une variable aléatoire.*

Alors, on définit la loi de probabilité de X *et on note* $\mathbb{P}_X$ *comme suit ; pour* B *in* $\mathcal{E}$, $\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B)$.

$\mathbb{P}_X$ *est parfois noté* $X * \mathbb{P}$.

$\mathbb{P}_X$ *est une mesure de probabilité de* $(E, \mathcal{E})$.

Definition A.4. (Variable aléatoire suivant une loi) Etant donné P une mesure de probabilité de $(E, \mathcal{E})$.

On dit qu'une variable aléatoire $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (E, \mathcal{E})$ suit la loi P et on notera, $X \sim P$, pour signifier que $\mathbb{P}_X = P$

Definition A.5 (Loi de probabilité uniforme discrète)**.** Soit $(E, \mathcal{E})$ un espace mesurable. étant donné S une partie finie de E ; on appelle loi uniforme discrète ou distribution uniforme sur S, la mesure $\mathbb{U}_S$ de $(E, \mathcal{E})$ est définit comme suit ;

$\mathbb{U}_S(\{x\}) = \frac{\mathbb{1}_S(x)}{card(S)}$.

C'est une mesure de probabilité.

Proposition A.2 (Calcul de l'esperance d'une indicatrice pour une distribution uniforme)**.** *Etant donnée S une partie finie de X, et P la distribution unfiorme sur S. On à pour toute partie mesurable A de X :* $\underset{P}{\mathbb{E}}(A) = \frac{card(A \cap S)}{card(S)}$

A.3 Suite de variables aléatoires

Definition A.6 (Variables aléatoires indépendantes et identiquement distribuées)**.** on dit que suite de variable aléatoire $(S_1, ..., S_n)$ est indépendante et identiquement distribué (on note i.i.d.) si et seulement,

1. Les $(S_i)_{1 \leq i \leq n}$ sont deux a deux indépendantes.
2. Pour tout $1 \leq i \leq n$, $\mathbb{E}(S_i) = \mathbb{E}(S_1)$

Proposition A.3 (Sur la Distribution Uniforme)**.** *Etant donnée S une partie finie de X, et P la distribution uniforme sur S. Quel que soit* $T = (T_i)_{1 \leq i \leq d}$ *une famille de d variable aléatoire i.i.d. suivant la loi P.*

On à l'égalité suivante ; $\mathbb{P}(T \in S^n) = 1$

Démonstration.

$$\begin{aligned}
\mathbb{P}(T \in S^n) &= \mathbb{P}(\{\omega \in \Omega \mid \forall 1 \leq i \leq n, T_i(\omega) \in S\}) \\
&= \mathbb{P}(\bigcap_{1 \leq i \leq n} \{\omega \in \Omega \mid T_i(\omega) \in S\}) \\
&= \mathbb{P}(\bigcap_{1 \leq i \leq n} T_i^{-1}(S)) \\
&= \prod_{1 \leq i \leq n} \mathbb{P}(T_i^{-1}(S)) \\
&= \prod_{1 \leq i \leq n} T_i * \mathbb{P}(S) \\
&= \prod_{1 \leq i \leq n} P(S) \\
&= \prod_{1 \leq i \leq n} 1 \\
&= 1
\end{aligned} \tag{1}$$

□

Annexe B Bon ordre, Ordinaux et cardinaux

B.1 Bon Ordre et Segment Initial

B.1.1 Bon Ordre

Definition B.1 (bon ordre)**.** Étant donné un ensemble E. on dit qu'une relation d'ordre $\leq$ sur E, est un bon ordre, si et seulement si,

Tout partie non vide S de E admet un plus petit élément pour $\leq$.

On dit alors que $(E, \leq)$ est un ensemble bien ordonné.

Definition B.2 (Morphisme d'ordre)**.** Étant donné $(E, \leq)$ et $(F, \preceq)$ deux ensemble ordonnés. On dit que $f : (E, \leq) \to (F, \preceq)$ est un morphisme d'ordre, si et seulement si ;

f est croissante, c'est à dire que, pour tout $x, y \in E$, on à
$x \leq y \Rightarrow f(x) \preceq f(y)$

Proposition B.1. *Étant donné $(E, \leq)$ un ensemble ordonné. Quel que soit $f : (E, \leq) \to (E, \leq)$ est un morphisme d'ordre,*

Pour tout $x \in E$, $x \leq f(x)$

Démonstration. Posons $W = \{x \in X \mid f(x) \leq x \quad et \quad x \neq f(x)\}$. Par l'absurde, supposons W non vide. Soit alors x_0 le plus petit élément de W.

Soit alors x un élément de E quelconque. Par minimalité de x_0, si on à la relation $x \leq x_0$; alors, $x \in W$ implique $x = x_0$. De fait, par contraposée, si $x \neq x_0$ alors $x \notin W$, c'est à dire que, $x \leq f(x)$.

De plus, si $x \leq x_0$ et $x \neq x_0$, alors f étant un automorphisme d'ordre, $f(x) \leq f(x_0)$ et $f(x) \neq f(x_0)$.

Mais, puisque $x_0 \in W$, d'une part on à $f(x_0) \leq x_0$, dès lors, $f(f(x_0)) \leq f(x_0)$, c'est à dire que $f(x_0) \in W$. D'autre part, $x_0 \in W$ indique $x_0 \neq f(x_0)$.

Donc on à, $f(x_0) \leq x_0$ et $f(x_0) \in W$, on en déduit que $f(x_0) = x_0$. On à donc une contradiction.

Donc par l'absurde on conclut que, W est vide. C'est à dire que, $E = W^c$, ie $\forall x \in E$, $x \leq f(x)$ □

B.1.2 Segment Initial

Definition B.3 (Segment initial)**.** Soit $(X, \leq)$ un ensemble ordonné. On dit qu'une partie $S \subseteq X$ est un segment initial, si et seulement si,

Pour tout élément x, y de X, si $y \in S$ et $x \leq y$, alors $x \in S$.

Definition B.4 (Segment des prédécesseur)**.** Soit $(X, \leq)$ un ensemble ordonné.
On pose $S_x = \{y \in S \mid y < x\}$

Proposition B.2 (Segment des prédécesseur est initial)**.** *Soit $(X, \leq)$ un ensemble bien ordonné.*

Quel que soit x dans X, S_x est un segment initial de $(X, \leq)$

Démonstration. en effet considérons $a \in S_x$ (c'est à dire que $a \leq x$), puis, y un élément de X tels que, $y \leq a$.

Par transitivité, on en déduit, $y \leq x$, ie. $y \in S_x$ $\square$

Proposition B.3 (Isomorphisme Segment initial)**.** *Soit $(X, \leq)$ un ensemble bien ordonné. W un segment initial de $(X, \leq)$. Et soit $f : X \to W$ une application de X dans W.*

Si f est un isomorphisme d'ordre, alors, $W = X$ et, pour tout $x \in X$, $f(x) = x$

Démonstration. On veut montrer que $W = X$, sachant que W est une partie de X, il nous suffira de montrer que, $X \subseteq W$.

Soit donc x un élément de X il s'agit de montrer que x est dans W. f est un morphisme d'ordre donc de X dans X donc on en déduit par la proposition $B.1$ que, $x \leq f(x)$.

Et, $f(x) \in W$, mais de plus, W est un segment initial. Donc de l'inégalité on déduit, $x \in W$.

D'où $X = W$

Et de plus, $f^{-1} : X \to X$ est un homomorphisme d'ordre. Donc quelque soit x dans X, $x \leq f^{-1}(x)$. Donc en appliquant f, par croissance de f ; $f(x) \leq x$ et d'autre part, $x \leq f(x)$. D'où, $x = f(x)$.

On à ainsi montrer que, f corresponds à la fonction identité. $\square$

Proposition B.4 (Comparaison de Segment initial)**.** *Soit $(X, \leq)$ un ensemble bien ordonné.*

Quel que soient V et W deux segments initiaux de $(X, \leq)$, on à $V \subseteq W$ ou $W \subseteq V$

Démonstration. Soient V et W deux segments initiaux. Supposons $\neg(W \subseteq V)$ dès lors, il existe w dans W, tel que $w \notin V$.

Alors nécessairement pour tout $v \in V$, $\neg(w \leq v)$, ie. $v \leq w$ et $w \neq v$.

Donc à fortiori, $v \leq w$, mais W est un segment initial, on en déduit donc que, $v \in W$. On conclut donc $V \subseteq W$.

Ainsi on à montrer que, $V \subseteq W$ ou $W \subseteq V$. $\square$

Proposition B.5 (Identification des segments initiaux)**.** *Soit $(X, <)$ un ensemble bien ordonnée, quel que soit W un segment initial de X ;*

$W = X$ ou, il existe x un élément de X tel que, $W = S_x$.

Démonstration. Supposons $W \neq X$, et considérons le minimum $min(X \setminus W) := x$. Montrons que $W = S_x$.

Soit $y \in W$, $y < x$ nécessairement ; sinon, $x \leq y$ et donc W étant un segment initial, $x \in W$. D'où, $y \in S_x$.

D'autre part, si $z \in S_x$, alors, $z < x$, si $z \in X \setminus W$, ceci contredit la minimalité de x, donc nécessairement, $z \in W$

On conclut $W = S_x$ $\square$

Proposition B.6. *Soit X un ensemble, $(W, \prec)$ un ensemble bien ordonnée.*

Si $f : X \to W$ est une bijection, en posant sur X la relation $<$ définit comme ; $\forall\ x, y \in X$, $x < y \Leftrightarrow f(x) \prec f(y)$

Alors, $<$ un bon ordre sur X, et de plus $(X, <)$ et $(W, \prec)$ sont isomorphes (par f).

B.2 Ordinaux, Classe Cardinale, Cardinal

B.2.1 Ordinaux

Definition B.5 (Ensemble Transitif)**.** On dit qu'un ensemble E est transitif, si et seulement si, quel que soit x un élément de X, $x \subseteq E$

On pourrait noter ; $\bigcup X \subseteq X$

Definition B.6 (Ordinal)**.** On dit qu'un ensemble α est un ordinal, si et seulement si,

1. α est un ensemble transitif
2. La relation d'appartenance $\in$ sur α est un bon ordre.

Proposition B.7 (Éléments d'un Ordinal)**.** *Quel que soit ω un ordinal, pour tout élément x de ω, x est un ordinal.*

Démonstration. Soit $x \in \omega$. Du fait que ω est un ensemble transitif, x est un partie de ω. Et puisque ω est un bien ordonnée pour $\in$; x est bien ordonnée pour $\in$.

Montrons que x est transitif ; soit y un élément de x, ie. $y \in x$. A fortiori donc, y est un élément de ω, donc par transitivité de ω ; y est une partie de ω. On veut montrer ; $y \subseteq x$, soit donc $a \in y$, puisque y est une partie alors, $a \in \omega$. Par transitivité de $\in$ on déduit, $a \in x$. D'où l'inclusion, $y \subseteq x$.

Donc x est un ensemble transitif, on à bien montrer que x est un ordinal. □

Proposition B.8 (Ordinal successeur)**.** *Étant donné α un ordinal, $\alpha \cup \{\alpha\}$ est aussi un ordinal.*

Démonstration. Montrons que $\alpha \cup \{\alpha\}$ est un ensemble transitif. Soit $y \in \alpha \cup \{\alpha\}$ on veut montrer que $y \subseteq \alpha \cup \{\alpha\}$.

Si $y \in \alpha$, alors nécessairement puisque α est un ensemble transitif ; $y \subseteq \omega$, et donc, $y \subseteq \alpha \cup \{\alpha\}$

Si $y = \alpha$ on à bien, $y \subseteq \alpha$ donc $y \subseteq \alpha \cup \{\alpha\}$.

Montrons que $\alpha \cup \{\alpha\}$ est bien ordonnée. Soit x et y deux élément distincts de $\alpha \cup \{\alpha\}$. On raisonne aussi par cas ; □

Definition B.7. (Ordinal Successeur) Etant donné α, l'ordinal $\alpha \cup \{\alpha\}$ est appelé successeur de α. Il est noté $\alpha + 1$.

Definition B.8. (Ordinal Limite) Un ordinal non vide, qui n'est successeur d'aucun autre ordinal est appelé ordinal limite.

Definition B.9. (Ordinal finie) Un ordinal α est dit fini, si, ni lui même, ni aucun de ses éléments n'est un ordinal limite.

Definition B.10. (Ordinal infinie) Un ordinal α est dit infini, si il n'est pas fini.

Notation B.1. (Omega dénombrable) Le plus petit ordinal infini sera noté ω.

Proposition B.9. *Soit α un ordinal. les assertions sont équivalentes ;*

1. *α est un ordinal fini*
2. *$\alpha \in \omega$*

Proposition B.10. *Soit α un ordinal, pour tout $\beta \in \alpha$; $S_\beta = \beta$*

Proposition B.11. *Soient α et β deux ordinaux, les assertions sont équivalentes ;*

1. $\alpha \subseteq \beta$
2. $\alpha \in \beta$ *ou* $\alpha = \beta$

Proposition B.12 (Trichotomie de la comparaison ordinale)**.** *Soient α et β deux ordinaux, une et une seule des trois assertions est vrai ;*

1. $\beta \in \alpha$
2. $\alpha \in \beta$
3. $\alpha = \beta$

B.2.2 Subpotence, Equipotence

Definition B.11 (Subpotence)**.** On dit qu'un ensemble X est subpotent à un ensemble Y, si et seulement si, il existe une injection $f : X \to Y$ de X vers Y.

On note alors $|X| \leq |Y|$

Definition B.12 (Equipotence)**.** On dit que deux ensembles X et Y sont équipotent, si et seulement si, il existe une bijection $f : X \to Y$ de X vers Y.

Theorem B.1 (Cantor-Bernstein)**.** *Étant donné X et Y deux ensembles si X est subpotent à Y, et Y est subpotent à X,*

Alors, X et Y sont équipotent.

Démonstration. Admis. □

B.2.3 Classe Cardinale

Definition B.13 (Classe Cardinale)**.** Étant donné X un ensemble, on appelle classe cardinale de X et on note $card(X)$, la classe des éléments équipotents à X.

Si λ est une classe cardinal, on dira que X est de cardinalité λ pour signifier, $X \in \lambda$

Notation B.2. La classe cardinal de ω est notée $\aleph_0$.

Definition B.14 (Ensemble Fini)**.** On dit qu'un ensemble est fini (ou est de cardinal fini), si il est équipotent à un ordinal fini.

Definition B.15 (Comparaison de classes cardinales)**.** Soient λ et μ deux classes cardinales.

On dit que λ est inférieur à μ et on notera $\lambda \leq \mu$, si et seulement si,

Il existe x un ensemble de cardinalité λ, et y un ensemble de cardinalité μ, tels que, x est subpotent à y.

Proposition B.13 (La comparaison des classes cardinales est une relation d'ordre)**.** *Soient λ et μ deux classes cardinales, on à les assertions suivantes ;*

1. $\lambda \leq \lambda$
2. $\lambda \leq \mu$ *et* $\mu \leq \lambda$, *implique* $\lambda = \mu$
3. $\lambda \leq mu$ *et* $\mu \leq \nu$, *implique* $\lambda \leq \nu$

Proposition B.14 (Subpotence et Produit Cartesien)**.** *Soient X, Y, Z et T des ensembles.*

Si X est subpotent à Z, et Y est subpotent à T, alors $X \times Y$ est subpotent à $Z \times T$.

Démonstration. Soient $f : X \to Z$ et $g : Y \to T$ deux injections. On pose alors $h : X \times Y \to Z \times T$ qui à (x, y) associe $(f(x), g(y))$.

Montrons que h est injective. Soient (x, y) et (u, v) deux couples de $X \times Y$ si on suppose $h(x, y) = h(u, v)$, alors on à $f(x) = f(u)$ donc par injectivité, $x = u$. Et on a $g(y) = g(v)$ et donc $y = v$. D'où $(x, y) = (u, v)$.

On viens d'expliciter une injection donc on en déduit que, $X \times Y$ est subpotent à $Z \times T$. □

Corollary B.1.1 (Equipotence et Produit Cartesien)**.** *Soient X, Y, Z et T des ensembles.*

Si X est equipotent à Z, et Y est equipotent à T, alors $X \times Y$ est equipotent à $Z \times T$.

Démonstration. On applique la proposition précédente, puis le théorème de Cantor-Bernstein. □

B.2.4 Cardinaux

Definition B.16 (Cardinal)**.** On dit qu'un ordinal $\aleph$ est un cardinal, si et seulement si, pour tout ordinal $\alpha \in \aleph$, α n'est pas équipotent à $\aleph$

Theorem B.2. *ω est un cardinal.*

Corollary B.2.1 (Ensemble Fini)**.** *Soit X un ensemble, les assertions sont équivalentes :*

1. *X est fini*
2. $card(X) < \aleph_0$

Démonstration. Si X est fini, cela signifie qu'il existe α un ordinal fini tel que, X et α sont équipotent. C'est à dire que $card(X) = card(\alpha)$.

α est un ordinal fini ceci équivaut à ; $\alpha \in \omega$. Par transitivité alors, $\alpha \subseteq \omega$. Donc α s'injecte dans ω, ie. $card(\alpha) \leq card(\omega)$.

Montrons que ω ne s'injecte pas dans α. En effet, $\alpha \in \omega$ et ω est un cardinal. Nécessairement donc ω ne s'injecte pas dans α, sinon on a leurs équipotence.

D'ou, $card(\alpha) < card(\omega)$, ie. $card(X) < \aleph_0$.

Réciproquement, si $card(X) < \aleph_0$. autrement dit, $card(X) < card(\omega)$. Remarquons que X est équipotent à un ordinal α, et donc $card(\alpha) = card(X)$. On à donc, $card(\alpha) < card(\omega)$: montrons que $\alpha \in \omega$. L'inégalité indique que, α s'injecte dans ω.

On à trois cas possible : $\alpha \in \omega$, $\omega \in \alpha$ ou $\alpha = \omega$.

L'égalité $\alpha = \omega$ des ordinaux, entrainerait l'égalité de leurs classes cardinal, donc on sais qu'elle n'est pas valide.

Si $\omega \in \alpha$ alors, ω est subpotent à α, et donc il en suit que α et ω sont équipotents, ie. ils ont la même classe cardinale, ceci contredirais l'inégalité stricte, donc nécessairement ce cas ne peut pas être vrai.

Donc par élimination des autres cas on à nécessairement, $\alpha \in \omega$. Ceci équivaut au fait que, α est un ordinal fini. Et puisque, $card(\alpha) = card(X)$ ceci signifie que, X est un ensemble fini. □

Proposition B.15 (Borne supérieur d'une famille de cardinaux)**.** *Étant donné A un ensemble de cardinaux; sup(A) est un cardinal.*

Démonstration. Posons $\alpha = sup(A)$. Soit alors β un ordinal strictement inférieur à α, ie. $\beta \in \alpha$.

Nécessairement, β ne majore donc pas la famille A. Dès lors, il existe γ un ordinal de A, tel que, $\beta < \gamma$, ie. $\beta \in \gamma$.

Et d'autre part comme γ appartient à A, on à : $\gamma \leq \alpha$ d'où, $card(\gamma) \leq card(\alpha)$.

Montrons $card(\beta) < card(\alpha)$. Si, α est subpotent à β, alors α et β sont équipotents. Mais γ est subpotent à α, ceci implique donc que, γ est subpotent à β. Mais $\beta \in \gamma$ et γ est un cardinal. Il y a contradiction.

Donc nécessairement, α n'est pas subpotent à β, et on conclut ainsi; $card(\beta) < card(\alpha)$.

Ceci montre que α est un cardinal. □

Proposition B.16 (Carré d'un cardinal infini)**.** *Étant donné λ un cardinal infini.*

$\lambda \times \lambda$ est équipotent à λ

Corollary B.2.2 (Puissance d'un cardinal infini)**.** *Étant donné λ un cardinal infini.*

quel que soit k un entier non nul, λ^k est équipotent à λ

Proposition B.17 (Cardinal d'union de produit, avec un ensemble infini)**.** *Étant donné X et Y deux ensembles non vides, dont l'un est infini.*

$card(X \times Y) = sup(card(X), card(Y))$

Démonstration. Posons $\lambda = sup(card(X), card(Y))$, par hypothèse, λ est infini.

Remarquons que, étant donné y un élément de Y (Y est supposé non vide); $\phi_y : X \to X \times Y$, $x \to (x, y)$ est une injection de X dans $X \times Y$. De fait; $card(X) \leq card(X \times Y)$

De même on montre que $card(Y) \leq card(X \times Y)$.

D'où $\lambda \leq card(X \times Y)$.

Ensuite, remarquons que $card(X) \leq \lambda$, et $card(Y) \leq \lambda$. On en déduit que, $card(X) \times card(Y) \leq \lambda \times \lambda$ c'est à dire que $card(X \times Y) \leq \lambda \times \lambda$

Mais par la proposition B.16, λ étant infini ; $\lambda \times \lambda$ est équipotent à λ. Mais l'inégalité précédente signifie que $X \times Y$ s'injecte dans $\lambda \times \lambda$

On en déduit donc que, $X \times Y$ s'injecte dans λ, c'est à dire, $card(X \times Y) \leq \lambda$

Des deux inégalités on conclut ; $card(X \times Y) = \lambda$

□

Corollary B.2.3 (Cardinal de la puissance d'un ensemble infini)**.** *Étant donné X un ensemble infini,*

Pour tout entier k non nul ; $card(X^k) = card(X)$

Démonstration. Par récurrence sur k. □

Proposition B.18 (Majoration du cardinal d'une union d'ensemble quelconque)**.** *Soient I un ensemble quelconque et $(X_i)_{i \in I}$ une famille d'ensemble, dès lors on à ;*

$$card(\bigcup_{i \in I} X_i) \leq card(I \times sup\{card(X_i) : i \in I\})$$

Démonstration. On pose $sup\{card(X_i) : i \in I\} = \lambda$ et $X = \bigcup_{i \in I} X_i$

Pour tout $x \in \bigcup_{i \in I} X_i$, c'est à dire qu'il existe $i \in I$ tel que, $x \in X_i$. Donc l'ensemble $I_x = \{i \in I \mid x \in X_i\}$ est non vide.

Par l'axiome du choix, il existe donc une fonction f qui à tout $x \in \bigcup_{i \in I} X_i$ associe $f(x)$ tel que $f(x) \in I_x$, c'est à dire que $x \in X_{f(x)}$.

Par ailleurs, pour tout $i \in I$ puisque $card(X_i) \leq \lambda$, il existe une injection de X_i dans λ, et donc l'ensemble $Inj(X_i) = \{g \mid g \text{ est une injection de } X_i \text{ dans } \lambda\}$ est non vide.

Par le choix encore, il existe donc g qui à tout élément $i \in I$ associe g_i un élément de $Inj(X_i)$ c'est dire que c'est une injection de X_i vers λ.

Posons alors, $\phi : X \to I \times \lambda$ qui à x associe $(f(x), g_{f(x)}(x))$. Montrons que ϕ est injective : en effet, si, $\phi(x) = \phi(y)$, alors, $f(x) = f(y)$, et donc il en suit par ailleurs que, $g_{f(x)}(x) = g_{f(y)}(x) = g_{f(y)}(y)$ et donc par l'injectivité de $g_{f(y)}$ on conclut ; $x = y$.

On à une injection donc la subpotence ; $card(X) \leq card(I \times \lambda)$ □

B.2.5 Type D'ordre

Theorem B.3. *Pour tout ensemble bien ordonnée $(E, <)$, Il existe un unique ordinal ω, tel que $(E, <)$ et $(\omega, \in)$ sont isomorphes.*

D'où la définition suivante ;

Definition B.17 (Type d'ordre)**.** Étant donné $(E, <)$ un ensemble bien ordonnée, on appelle type d'ordre de $(E, <)$, l'unique ordinal isomorphe à $(E, <)$.

Theorem B.4 (Séparation cardinale)**.** *Soient* $(X, <)$ *un ensemble ordonné, de type d'ordre* α*, tel que* α *est un cardinal.*

Alors pour tout segment initial propre S *de* X *de type d'ordre* β*,* $card(S) < card(X)$

Démonstration. Par l'absurde supposons $card(S) = card(X)$ alors il en suit que, $card(\beta) = card(\alpha)$. On va raisonner par trichotomie de la comparaison ordinal.

Si, $\beta \in \alpha$, remarquons que α est un cardinal, et que par l'hypothèse β et α sont équipotents. Par la definition de cardinal, ces assertions sont incompatibles, on à une contradiction.

Si, $\alpha = \beta$, alors, S est de type d'ordre α, et ayant le même type d'ordre, S et X sont isomorphes. Or S est un segment initial de X, donc en vertu de la proposition B.3, on en déduit, $X = S$, mais ceci viens contredire le fait que S est un segment initial propre, ie. distinct de X.

Autrement si $\alpha \in \beta$ alors à fortiori, $\alpha \subseteq \beta$.

Considérons $f : X \to \alpha$ l'isomorphisme de X vers α. On peut appliquer f à S, et étant un isomorphisme celui ci conservera les segments initiaux, donc $f(S)$ est un segment initial de α et donc $f(S) \subseteq \alpha$. Mais par unicité du type d'ordre, $f(S) = \beta$, d'ou, $\beta \subseteq \alpha$.

Mais, $\alpha \in \beta$, donc ceci entraine, $\alpha \in \alpha$, ce qui est impossible, puisque α est un ordinal, et donc $\in$ est un ordre strict sur α. □

Theorem B.5 (Zermelo)**.** *Tout ensemble peut être muni d'un bon ordre*

Démonstration. Admis. Par ailleurs ce théorème est équivalent à l'axiome du choix qu'on admet ici. □

Corollary B.5.1. *Tout ensemble est équipotent à un ordinal.*

Démonstration. C'est une conséquence du théorème de Zermelo, puisqu'un isomorphisme est à fortiori une bijection. □

Corollary B.5.2. *Toute classe cardinal contient au moins un ordinal.*

Démonstration. Considérons λ une classe cardinale, par définition cela signifie qu'il existe un ensemble X tel que, $card(X) = \lambda$.

Mais, par le théorème de Zermelo, il existe un ordinal α isomorphe à X, alors à fortiori, α est équipotent à X, donc α est dans la classe cardinale de X, c'est à dire λ. □

Corollary B.5.3. *Toute classe cardinal contient un et un seul cardinal.*

Démonstration. Soient λ une classe cardinale. montrons que λ contient un et un seul cardinal.

Unicité. Soient α et β deux cardinaux dans λ, ceci signifie que α et β sont équipotents. On va raisonner par trichotomie pour la comparaison ordinal.

Si, $\alpha \in \beta$, alors puisque α est équipotent à β, ceci viens contredire le fait que β est un cardinal. Le cas $\beta \in \alpha$ est symétrique.

Nécessairement on à donc $\alpha = \beta$.

Existence. Par l'absurde, supposons qu'aucun cardinal n'est de cardinalité λ. Soit alors α un ordinal de cardinalité λ (Il en existe bien de par le corollaire B.5.2). Nécessairement, α n'est pas un cardinal, il existe donc, $\beta \in \alpha$ tel que, α et β sont équipotents.

Posons alors $\Lambda = \{\beta \in \alpha : \alpha$ et β sont équipotents $\}$. Λ est non vide et est une partie de α, α étant bien ordonné, on en déduit que Λ à un plus petit élément.

Soit γ le plus petit éléments de Λ, remarquons que γ est équipotent à α, de fait, γ est de cardinalité λ. Nécessairement alors, γ n'est pas un cardinal, ie. il existe $\beta \in \gamma$ tel que β est équipotent à γ, mais alors, β est équipotent à α. Et donc $\beta \in \Lambda$, mais $\beta \in \gamma$, donc ceci contredit la minimalité de γ.

Nécessairement donc il existe un cardinal de cardinalité λ. □

Corollary B.5.4. *Pour tout ensemble X de classe cardinale $card(X)$. α un (le seul) cardinal de $card(X)$*

On peut munir X d'un bon ordre $<$ tel que, $(X, <)$ et $(\alpha, \in)$ sont isomorphes.

Démonstration. En effet, $card(X) = card(\alpha)$, c'est à dire que, X et α sont équipotents, par la proposition B.6 on en déduit que, on peut construire un ordre sur X tel que, X et α sont isomorphes. □

B.3 Alephs et Omegas

Theorem B.6 (de Cantor)**.** *Pour tout classe cardinale* λ *;* $\lambda < 2^{\lambda}$

Corollary B.6.1 (Tout cardinal peut être majoré par un autre)**.** *Pour tout cardinal* α *il existe un ordinal* β *tel que,* $card(\alpha) < card(\beta)$

Proposition B.19 (Ordinal successeur)**.** *Soit* α *un cardinal, on appelle cardinal successeur de* α*, et on note* α^{+} *le plus petit cardinal supérieur à* α*.*

Et si λ *est une classe cardinale et* α *est un ordinal dans* λ*, on note* λ^{+} *la classe cardinale de* α^{+}*.*

On à tendance à ne pas distinguer les classes cardinales et les cardinaux.

Definition B.18. (Omegas) On définit la fonction ω de la classe des ordinaux dans la classe des cardinaux comme suit :

1. $\omega_0 = \omega$ c'est à dire le plus petit ordinal infini.
2. Si $\beta = \alpha + 1$ est un ordinal successeur, $\omega_\beta = (\omega_\alpha)^{+}$
3. si α est un ordinal limite, $\omega_\alpha = sup\{\omega_\beta; \beta \in \alpha\}$

Definition B.19. (Alephs) On définit la fonction $\aleph$ qui à un ordinal associe une classe cardinale comme suit :

Si α est un ordinal $\aleph_\alpha$ corresponds à la classe cardinale de ω_α ; $\aleph_\alpha = card(\omega_\alpha)$

Annexe C Théorie des Modèles

C.1 Syntaxe

Definition C.1 (Langage du premier ordre)**.** On appelle langage du premier ordre, un ensemble $\mathfrak{L}$ de symbole ;

- $\mathcal{V}$ commun à tous les langages est un ensemble infini dénombrable dont les élément sont appelés variables.
- des symboles connecteurs du calcul propositionnel $\{\neg,), (, \wedge, \vee, \Rightarrow, \Leftrightarrow\}$.
- d'un quantificateur universelle $\forall$ et d'un quantificateur existentiel $\exists$.

Ces ensembles sont communs à tous le langages du premier ordre. Les ensembles suivants en revanches sont ceux susceptibles de changer ;

- D'un ensemble $\mathcal{C}$ donc les symboles sont appelés symboles de constantes.
- Dune suite d'ensembles $(\mathfrak{F}_n)n \in \mathbb{N}^*$ deux a deux disjoints, de symboles dit de fonctions et si $f \in \mathfrak{F}_\mathbf{n}$ on dit que f est d'arité n.
- Dune suite d'ensembles $(\mathfrak{R}_n)n \in \mathbb{N}^*$ deux a deux disjoints, de symboles dit de prédicats ou relationnels, et si $R \in \mathfrak{F}_\mathbf{n}$ on dit que R est d'arité n.

Pour se donner un langage il suffira donc de se donner $\mathcal{C}$, $\bigcup\limits_{n\in\mathbb{N}^*} \mathfrak{F}_n$ et $\bigcup\limits_{n\in\mathbb{N}^*} \mathfrak{R}_n$

Definition C.2 ($\mathfrak{L}$-terme)**.** Soit $\mathfrak{L}$ un langage du premier ordre, on définit l'ensemble $\mathcal{T}(\mathfrak{L})$ des termes de $\mathfrak{L}$ comme le plus petit ensemble ;

- Qui contient tous les symboles de variables et constantes (ie. ce sont des termes).
- Pour tout entier n non nul, pour tout $f \in \mathfrak{F}_n$, $\mathcal{T}(\mathfrak{L})$ est stable par $(m_1, ...m_n) \mapsto f(m_1 m_2 ... m_n)$.

Par défaut maintenant, $\mathfrak{L}$ sera un langage du premier ordre.

Definition C.3 ($\mathfrak{L}$-formule atomique)**.** On dit qu'un mot sur $\mathfrak{L}$ est une formule atomique, si il existe un entier $n \in \mathbb{N}^*$, un symbole de relation n-aire R, et n termes $t_1, ..., t_n$ telles que ;

$$M = R(t_1 ... t_n)$$

Definition C.4 ($\mathfrak{L}$-formule)**.** On dit définit l'ensemble des formules $\mathcal{F}(\mathfrak{L})$ sur $\mathfrak{L}$ comme le plus petit ensemble des mots sur $\mathfrak{L}$ tels que ;

- Il contient les formules atomiques sur $\mathfrak{L}$
- Pour toute formule M et N, pour tout symbole de variable v ; $\neg M$ $(M \vee N)$, $(M \wedge N)$, $(M \Rightarrow N)$, $(M \Leftrightarrow N)$, $\forall v M$ et $\exists v M$ sont aussi des formules.

Definition C.5 (Variable libre, variable lié, énoncé)**.** Soit x une variable.

- Étant donné ϕ une formule de $\mathfrak{L}$; les occurrences de x dans $\forall x \phi$, $\exists x \phi$ sont dites liées, respectivement au quantificateur $\forall$ et $\exists$.
- Étant donné ϕ une formule de $\mathfrak{L}$; les occurrences de x qui ne sont liées à aucun quantificateur sont dites libres
- On appelle énoncé, ou formule close, toute formule dont toutes les occurrences de variables sont liées.

C.2 Semantique, Structure

Definition C.6 (Structure)**.** Une structure $\mathfrak{M}$ est la donnée d'un ensemble non vide M parfois nommé univers muni ;

- d'une famille $(c_i^{\mathfrak{M}})_{i\in I}$, de constantes telles que pour tout i, $c_i^{\mathfrak{M}} \in M$.
- d'une famille $(f_j^{\mathfrak{M}})_{j\in J}$, de fonctions ou pour tout j, il existe un entier n_j tel que $f_j^{\mathfrak{M}}$ est une fonction totale de M^{n_j} dans M.
- d'une famille $(R_k^{\mathfrak{M}})_{k\in K}$, de relations ou pour tout k, il existe un entier m_k tel que $R_k^{\mathfrak{M}}$ est un sous ensemble de M^{m_k}, donc une relation sur M d'arité m_k.
- Et munit de la relation d'égalité, la diagonale de M^2.

Definition C.7 (Langage associé)**.** On dit qu'un language du premier ordre $\mathfrak{L}$ est associé à une structure $\mathfrak{M}$ si ;

- pour tout $(c_i^{\mathfrak{M}})$, $\mathfrak{L}$ contient un symbole de constante c_i. Et ce sont ces seuls symboles de constantes.
- pour tout $(f_j^{\mathfrak{M}})$, $\mathfrak{L}$ contient un symbole de fonction f_j d'arité n_j. Et ce sont ces seuls symboles de fonctions.
- pour chaque $(R_k^{\mathfrak{M}})$, $\mathfrak{L}$contient un symbole de relation R_k d'arité m_k. Et ce sont ces seuls symboles de relations.

Et étant donné un symbole quelconque ϕ de $\mathfrak{L}$, on dit que $\phi^{\mathfrak{M}}$ est l'interprétation de du symbole ϕ dans la structure $\mathfrak{M}$.

On dira aussi que $\mathfrak{M}$ est une $\mathfrak{L}$-structure.

Definition C.8 (Interprétation des termes)**.** Soit $\mathfrak{M}$ est une $\mathfrak{L}$-structure. Étant donné t un terme de $\mathfrak{L}$, on note $t(\overline{x})$ pour signifier que, seuls les variables du vecteur $\overline{x}$ apparaissent dans t.

Étant donné $\overline{m}$ un vecteur d'éléments de M de même longueur que $\overline{x}$. On obtient $t(\overline{m})$ en substituant m_i à chaque occurrence de x_i.

Alors on définit l'interprétation $t^{\mathfrak{M}}(\overline{m})$ de $t(\overline{m})$ par induction comme suit ;

- L'interprétation d'une constante c est $c^{\mathfrak{M}}$
- L'interprétation d'un paramètre m est m
- L'interprétation, étant donné f une fonction n-aire et $t_1, ..., t_n$ des termes, l'interpretation $f(t_1, ..., t_n)^{\mathfrak{M}}(\overline{m})$ est définit comme $f^{\mathfrak{M}}(t_1^{\mathfrak{M}}(\overline{m}), ..., t_n^{\mathfrak{M}}(\overline{m}))$

Definition C.9 (Satisfaction d'un formule)**.** Soit $\mathfrak{M}$ est une $\mathfrak{L}$-structure. Étant donné ϕ une formule de $\mathfrak{L}$, on note $\phi(\overline{x})$ pour signifier que, seuls les variables du vecteur $\overline{x}$ apparaissent dans ϕ.

Étant donné $\overline{m}$ un vecteur d'éléments de M de même longueur que $\overline{x}$. On obtient $\phi(\overline{m})$ en substituant m_i à chaque occurrence de x_i.

Alors toujours par induction on définit la satisfaction de $\phi(\overline{m})$ dans $\mathfrak{M}$ que l'on notera $\mathfrak{M} \vDash \phi(\overline{m})$, comme suit ;

- $\mathfrak{M} \vDash R(t_1, ..., t_n)(\overline{m})$ si et seulement si, $(t_1^{\mathfrak{M}}(\overline{m}), ..., t_n^{\mathfrak{M}}(\overline{m})) \in R^{\mathfrak{M}}$
- $\mathfrak{M} \vDash \neg\phi(\overline{m})$ si et seulement si, $\mathfrak{M} \nvDash \phi(\overline{m})$
- $\mathfrak{M} \vDash (\phi_1 \wedge \phi_2)(\overline{m})$ si et seulement si, $\mathfrak{M} \vDash \phi_1(\overline{m})$ et $\mathfrak{M} \vDash \phi_2(\overline{m})$
- $\mathfrak{M} \vDash (\phi_1 \vee \phi_2)(\overline{m})$ si et seulement si, $\mathfrak{M} \vDash \phi_1(\overline{m})$ ou $\mathfrak{M} \vDash \phi_2(\overline{m})$
- $\mathfrak{M} \vDash \forall x\phi(x, \overline{m})$ si et seulement si, Pour tout $a \in M$, $\mathfrak{M} \nvDash \phi(a, \overline{m})$

- $\mathfrak{M} \models \exists x \phi(x, \overline{m})$ si et seulement si, Il existe $a \in M$, $\mathfrak{M} \not\models \phi(a, \overline{m})$

On dira pour signifier $\mathfrak{M} \models \phi(\overline{m})$, que, $\phi(\overline{m})$ est satisfaite dans $\mathfrak{M}$, que $\mathfrak{M}$ satisfait $\phi(\overline{m})$ ou bien encore que $\overline{m}$ satisfait $\phi(\overline{x})$ dans $\mathfrak{M}$.

C.3 Théorie, Modèle d'une théorie

Definition C.10 (Théorie)**.** On appelle théorie, tout ensemble T de formule close.

Definition C.11 (Modèle)**.** Etant donné T une théorie, on dit que $\mathfrak{M}$, est un modèle de T, si et seulement si, toutes les formules de T sont satisfaites dans $\mathfrak{M}$.

On note $\mathfrak{M} \models T$.

Definition C.12 (Conséquence, formule prouvable)**.** Soit ϕ une $\mathfrak{L}$-formule, et T une théorie.

On dit que ϕ est une conséquence sémantique de T, si et seulement si, tout modèle de T est satisfait ϕ.

On note alors $T \models \phi$.

Definition C.13 (Consistance)**.** On dit qu'une théorie T est consistante ou non contradictoire, si elle admet un modèle.

Sinon on dira qu'elle est inconsistante ou contradictoire.

C.4 Sous structure

Pour la partie sur la dimension sans doute qu'on déplacera cette sous partie

Definition C.14 (Sous-structure)**.** On dit que, une $\mathfrak{L}$ structure $\mathfrak{N}$ est une sous structure $\mathfrak{M}$, si et seulement si ;

- L'univers N de $\mathfrak{N}$, n'est pas vide, et $N \subseteq M$.
- Pour tout relation S de $\mathfrak{N}$ d'arité n, il existe R une relation de $\mathfrak{M}$ telle que, S corresponds a la restriction de R à l'univers N, ie. $S = R \cap N^n$.
- Pour toute fonction g de $\mathfrak{N}$ d'arité n, il existe g une fonction de $\mathfrak{M}$ telle que, g corresponds a la restriction de f à l'univers N, ie. $g = f_{|N}$.
- Toute constante de $\mathfrak{N}$ est aussi une constante de $\mathfrak{M}$.

Dans ce cas on note, $\mathfrak{N} \subseteq \mathfrak{M}$. On dit aussi que, $\mathfrak{M}$ est une extension de $\mathfrak{N}$.

Definition C.15 (Formule Borné)**.**

Definition C.16 (Propriété)**.**

Definition C.17 (Propriété à caractère finie)**.**

Références

[1] Shai Ben-David , Pavel Hrubeš , Shay Moran , Amir Shpilka and Amir Yehudayoff *Learnability can be undecidable* Link Article, Nature Machine Intelligence, 2019.

[2] Shai Ben-David, Pavel Hrubeš, Shay Moran, Amir Shpilka, and Amir Yehudayoff *On a learning problem that is independent of the set theory ZFC axioms* Link Article, 2019.

[3] René Cori, Daniel Lascar *Logique Mathématique volume I et II.* Dunod, 2003.

[4] Itaï Ben Yaacov, Thomas Blossier, Julien Melleray *Introduction à la Logique Mathématique* Link Note de cours, 2011.

[5] Michael J. Kearns, Umesh V. Vazirani *An Introduction to Computational Learning Theory* The MIT Press, Massachusetts,1994.

[6] Nick Littlestone, Manfred K. Warmuth *Relating Data Compression and Learnability* Link Technical Report, University of California, 1986.

[7] Shay Moran, Amir Yehudayoff *Sample compression schemes for VC classes* Link article, 2015.

[8] K. P. Hart *Machine learning and the Continuum Hypothesis* Link Article, Faculteit EWI, 2019.

[9] Thomas Blossier, Julien Melleray, Frank Wagner *Introduction à la Logique Mathématique, seconde partie : Théorie des modèles* Link Note de cours.

[10] Julien MELLERAY *Les axiomes de Zermelo-Fraenkel* Link Lyon.

[11] François Schwarzentruber *Logique, Préparation à l'option informatique de l'agrégation de mathématiques* Link ENS Rennes.

[12] Delia KESNER *Règles systemes LK, G, et DNprop* Link

[13] Patrick Dehornoy *Progrès récent sur l'hypothèse du continu (d'après Woodin)* Link article, 2004.

[14] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler et Manfred K. Warmuth *Learnability and the Vapnik-Chervonenkis Dimension* Link article, 1989.