Sur la consistence des forêts aléatoires

Analyse d'article

Dalmard Alban, Mirone Jeanne et Royer Jules

Mines Paris PSL

data.sophia@minesparis.psl.eu

26 mars 2025
Présentations par les élèves

Table des matières

1. Introduction
2. Classifieurs de vote et de moyenne
3. Forêts Aléatoires
4. Consistence par Aggregation
5. Bootstrap Aggregation
6. Arbre à croissance gourmande
7. Conclusion

1 Introduction

2 Classifieurs de vote et de moyenne

3 Forêts Aléatoires

4 Consistence par Aggregation

5 Bootstrap Aggregation

6 Arbre à croissance gourmande

7 Conclusion

Introduction : Les forêts aléatoires

- Problèmes de classification
- Méthode *ensembliste* : agrège prédictions de plusieurs classifieurs de base (ici des arbres de décision). Principe de la "Sagesse des foules".
- Méthodes d'agrégation : moyenne, mode (vote) de la cellule contenant l'observation X.

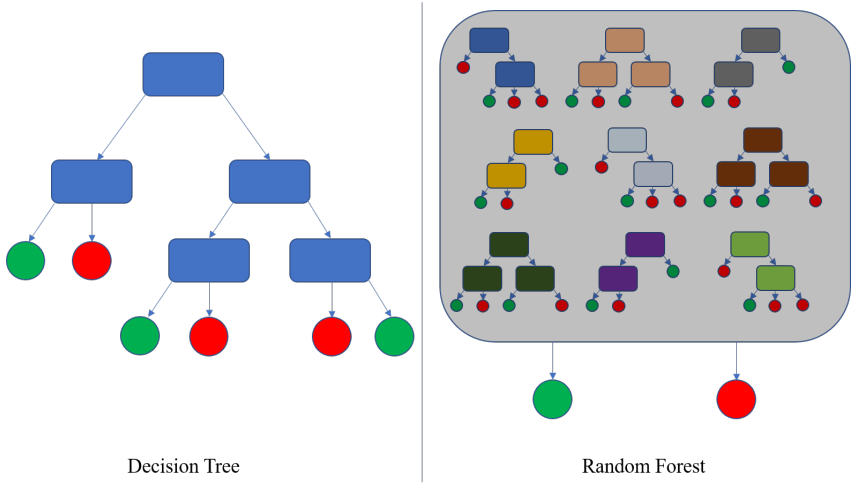


Figure 1 – Arbre de décision et forêt aléatoire.

- Introduction d'aléatoire dans la forêt : arbre de décision ou jeu d'entraînement construit aléatoirement (*bagging*).
- Intérêts principaux : bonnes performances, forêt aléatoire de classifieurs de base inconsistants peut être consistante

1 Introduction

2 Classifieurs de vote et de moyenne

3 Forêts Aléatoires

4 Consistence par Aggregation

5 Bootstrap Aggregation

6 Arbre à croissance gourmande

7 Conclusion

Classificateurs de vote et de moyenne

- $(X, Y), (X_1, Y_1), ..., (X_n, Y_n)$ iid $X \in \mathbb{R}^d$ et $Y \in \{0, 1\}$
- Probabilité à postériori $\eta : \mathbb{R}^d \to [0, 1]$ defini par : $\eta(x) = P\{Y = 1 | X = x\}$
- $(X_1, Y_1), ..., (X_n, Y_n)$ données d'entrainement D_n
- Probabilité d'erreur : $L(g_n) = P(X, Y)\{g_n(X, D_n) \neq Y\}$
- Classifieur de Bayes : $g^*(x) = 1\{\eta(x) \geq 1/2\}$
- Un classifieur est consistant pour (X, Y) si $\lim_{n\to\infty} L(g_n) \to L^*$

Classifieurs de vote et de moyenne

- Classifieur aléatoire : $g_n(x, Z, D_n)$
- Classifieur de vote : $g_n^{(m)}(x, Z_m, D_n) = \begin{cases} 1 & \text{si } \frac{1}{m}\sum_{j=1}^{m} g_n^{(m)}(x, Z_j, D_n) \geq \frac{1}{2}, \\ 0 & \text{sinon.} \end{cases}$
- Classifieur moyenné : $\lim_{m\to\infty} g_n^{(m)}(x, Z_m, D_n) = \overline{g_n} = \mathbf{1}_{\{\mathbb{E}_Z[g_n(x,Z)]\}\geq 1/2}$

Classifieurs de vote et de moyenne

Proposition : Consistance d'un classificateur aléatoire moyenné

Si les classifieurs aléatoiresg_n son consistant pour une distribution (X, Y) données, alors le classifieur de vote $g_n(m)$ et le classifieur moyenné $\overline{g_n}$ sont consistants.

Proof Consistency of $\{g_n\}$ is equivalent to saying that $\mathbb{E}L(g_n) = \mathbb{P}\{g_n(X,Z) \neq Y\} \to L^*$. In fact, since $\mathbb{P}\{g_n(X,Z) \neq Y | X = x\} \geq \mathbb{P}\{g^*(X) \neq Y | X = x\}$ for all $x \in \mathbb{R}^d$, consistency of $\{g_n\}$ means that for μ-almost all x,

$$\mathbb{P}\{g_n(X,Z) \neq Y | X = x\} \to \mathbb{P}\{g^*(X) \neq Y | X = x\} = \min(\eta(x), 1-\eta(x)) \, .$$

Without loss of generality, assume that $\eta(x) > 1/2$. (In the case of $\eta(x) = 1/2$ any classifier has a conditional probability of error $1/2$ and there is nothing to prove.) Then $\mathbb{P}\{g_n(X,Z) \neq Y | X = x\} = (2\eta(x) - 1)\mathbb{P}\{g_n(x,Z) = 0\} + 1 - \eta(x)$, and by consistency we have $\mathbb{P}\{g_n(x,Z) = 0\} \to 0$.

To prove consistency of the voting classifier $g_n^{(m)}$, it suffices to show that $\mathbb{P}\{g_n^{(m)}(x,Z^m) = 0\} \to 0$ for μ-almost all x for which $\eta(x) > 1/2$. However,

$$\begin{aligned}
\mathbb{P}\{g_n^{(m)}(x,Z^m) = 0\} &= \mathbb{P}\left\{ (1/m) \sum_{j=1}^{m} \mathbb{1}_{\{g_n(x,Z_j)=0\}} > 1/2 \right\} \\
&\leq 2\mathbb{E}\left[(1/m) \sum_{j=1}^{m} \mathbb{1}_{\{g_n(x,Z_j)=0\}} \right] \\
&\qquad \text{(by Markov's inequality)} \\
&= 2\mathbb{P}\{g_n(x,Z) = 0\} \to 0 \, .
\end{aligned}$$

Consistency of the averaged classifier is proved by a similar argument.

□

Figure 2 – Démonstration proposition 1

1 Introduction

2 Classifieurs de vote et de moyenne

3 Forêts Aléatoires

4 Consistence par Aggregation

5 Bootstrap Aggregation

6 Arbre à croissance gourmande

7 Conclusion

Définition

- Forêt aléatoire définie comme classifieur moyennant une infinité d'arbres construits aléatoirement $\overline{g_n} = \mathbf{1}_{\{\mathbb{E}_Z[g_n(x,Z)] \geq 1/2\}}$
- Aléatoire : les Z_i sont id. distribués, et indépendants une fois X, Y et D_n fixés. Régit la construction du classifieur de base. Choix cellule à couper, direction, puis abscisse de coupe.
- 2 modèles simples d'arbre de partition comme classifieurs de base : *forêt purement aléatoire* et *forêt invariante par échelle*. $\rightarrow$ Illustration de la démarche de démonstration de la consistance d'une forêt aléatoire.

Forêt purement aléatoire

Hypothèse : $X \in [0,1]^d$p.s information sur l'étendue spatiale des features.

Construction aléatoire de l'arbre de décision $g_n(x, Z)$:

- A chaque étape, les *feuilles* partitionnent $[0,1]^d$.
- k coupes de noeud/cellule : Choix aléatoires et uniformes. Une feuille, une direction de coupe de la cellule parmi $x^{(1)}, ..., x^{(d)}$, et une abscisse $\lambda \in [0,1]$ de coupe.
- k peut dépendre de n.

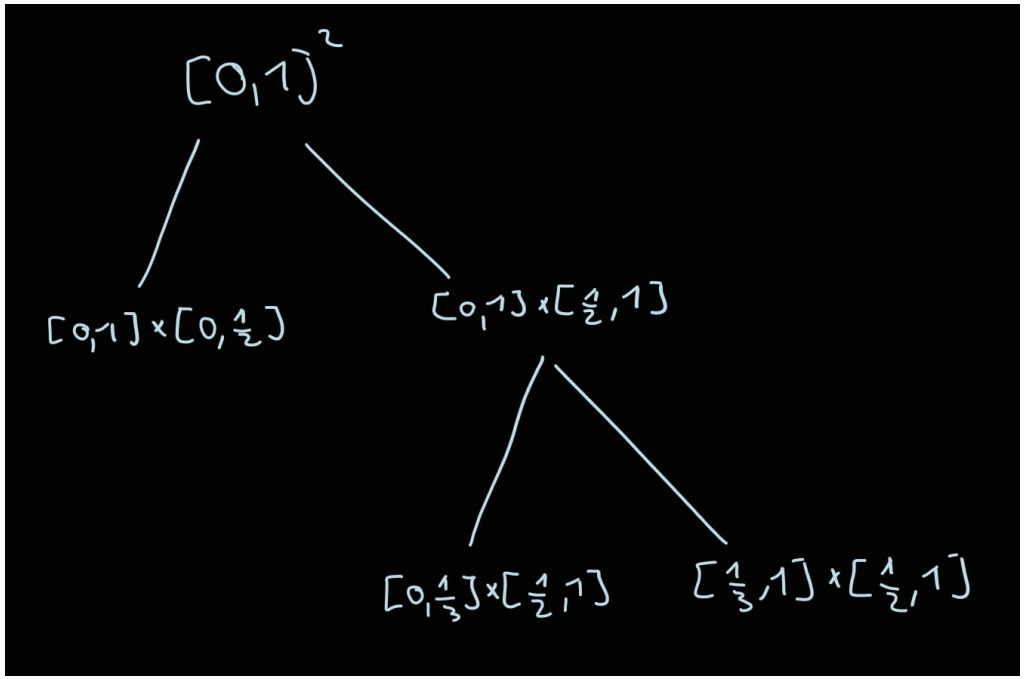


Figure 3 – Arbre de partition, $k = 2$ coupes, selon $x^{(2)}$ puis $x^{(1)}$.

Théorème : Consistance de la forêt purement aléatoire

Si $[0,1]^d$ est le support de X, alors la forêt purement aléatoire $\overline{g_n}$ est consistante si $k_n \to \infty$ et $k_n/n \to \infty$ quand $n \to \infty$.

- Conditions similaires à celle de COVERT et HART : $\|X_{(k_m)}(X) - X\| \xrightarrow{\text{p.s}} 0$

Proof By Proposition 1 it suffices to prove consistency of the randomized base tree classifier g_n. To this end, we recall a general consistency theorem for partitioning classifiers proved in (Devroye, Györfi, and Lugosi, 1996, Theorem 6.1). According to this theorem, g_n is consistent if both $\text{diam}(A_n(X,Z)) \to 0$ in probability and $N_n(X,Z) \to \infty$ in probability, where $A_n(x,Z)$ is the rectangular cell of the random partition containing x and

$$N_n(x,Z) = \sum_{i=1}^{n} \mathbb{1}_{\{X_i \in A_n(x,Z)\}}$$

is the number of data points falling in the same cell as x.

Preuve : Consistance de la forêt purement aléatoire

- Théorème de partitionnement garantit $\mathbb{E}_{P^m}[L(g_m)] \xrightarrow{n\to\infty} L^*$, ie convergence L_1 donc aussi en proba.

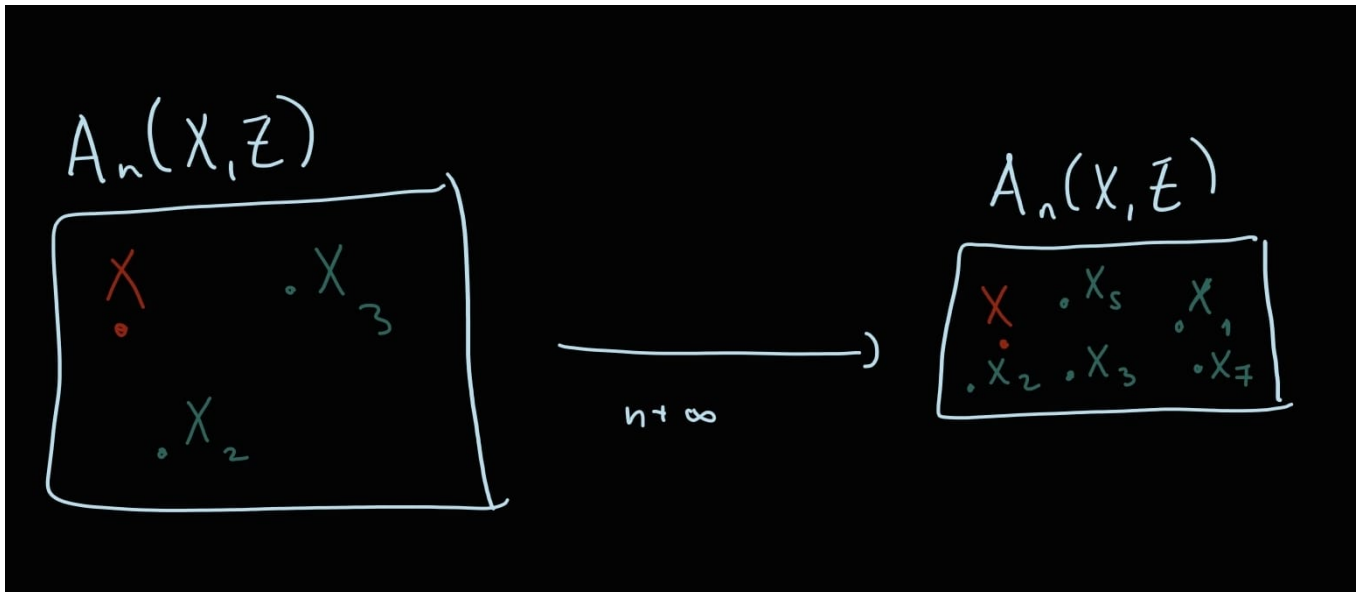


Figure 4 – Hypothèses de concentration : compromis entre *diversité* (avoir assez de voisins pour réduire la variance du vote) et *localité* (dans une zone assez petite pour prendre en compte des variations fines de la distribution).

First we show that $N_n(X,Z) \to \infty$ in probability. Consider the random tree partition defined by Z. Observe that the partition has $k+1$ rectangular cells, say $A_1,\ldots,A_{k+1}$. Let $N_1,\ldots,N_{k+1}$ denote the number of points of $X,X_1,\ldots,X_n$ falling in these $k+1$ cells. Let $S=\{X,X_1,\ldots,X_n\}$ denote the set of positions of these $n+1$ points. Since these points are independent and identically distributed, fixing the *set* S (but not the order of the points) and Z, the conditional probability that X falls in the i-th cell equals $N_i/(n+1)$. Thus, for every fixed $t>0$,

$$\begin{aligned}\mathbb{P}\{N_n(X,Z)<t\} &= \mathbb{E}\left[\mathbb{P}\{N_n(X,Z)<t|S,Z\}\right] \\ &= \mathbb{E}\left[\sum_{i:N_i<t}\frac{N_i}{n+1}\right] \leq (t-1)\frac{k+1}{n+1}\end{aligned}$$

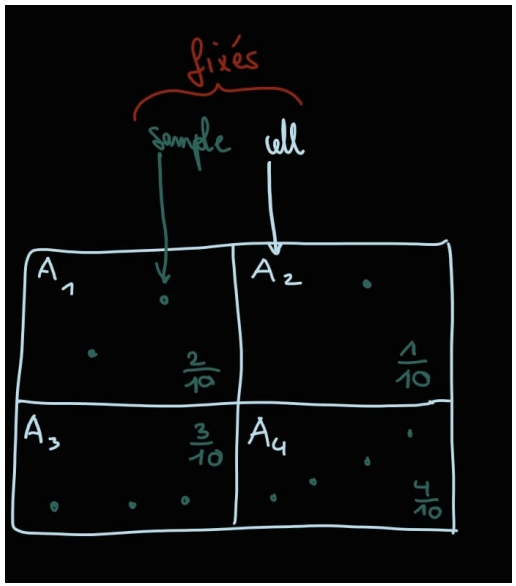


Figure 5 – Proportions d'échantillons dans chaque cellule, à S, Z fixés.

It remains to show that $\text{diam}(A_n(X,Z)) \to 0$ in probability. To this aim, let $V_n = V_n(x,Z)$ be the size of the first dimension of the rectangle containing x. Let $T_n = T_n(x,Z)$ be the number of times that the box containing x is split when we construct the random tree partition.

Let K_n be binomial $(T_n, 1/d)$, representing the number of times the box containing x is split along the first coordinate.

Clearly, it suffices to show that $V_n(x,Z) \to 0$ in probability for μ-almost all x, so it is enough to show that for all x, $\mathbb{E}[V_n(x,Z)] \to 0$. Observe that if $U_1, U_2, \ldots$ are independent uniform $[0,1]$, then

$$\mathbb{E}[V_n(x,Z)] \quad \leq \quad \mathbb{E}\left[\mathbb{E}\left[\prod_{i=1}^{K_n} \max(U_i, 1-U_i) \middle| K_n\right]\right]$$

- $\text{diam}(A)^2 \leq \sum_{i=1}^{d} V_i^2$ où V_i est coordonnée du diamètre selon l'axe $x^{(i)}$
- $\mathbb{E}_Z[V_n(x,Z)] \to 0 \Rightarrow V_n(x,Z) \xrightarrow{\mathbb{P}_Z} 0$
- conditionnement selon le nombre de coupes selon l'axe 1.

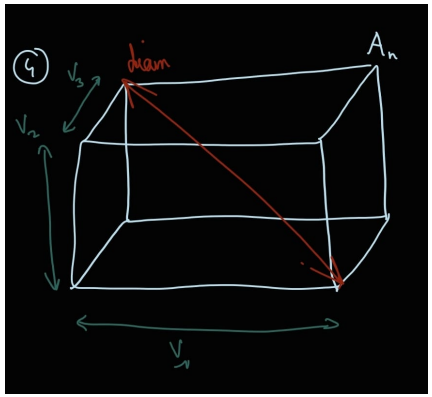


Figure 6 – Majoration du diamètre par une somme de côtés.

$$
\begin{aligned}
\mathbb{E}[V_n(x,Z)] &\leq \mathbb{E}\left[\mathbb{E}\left[\prod_{i=1}^{K_n}\max(U_i,1-U_i)\middle| K_n\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\max(U_1,1-U_1)\right]^{K_n}\right] \\
&= \mathbb{E}\left[(3/4)^{K_n}\right] \\
&= \mathbb{E}\left[\left(1-\frac{1}{d}+\frac{3}{4d}\right)^{T_n}\right] \\
&= \mathbb{E}\left[\left(1-\frac{1}{4d}\right)^{T_n}\right].
\end{aligned}
$$

? ? $K_n \rightsquigarrow \mathcal{B}(T_n, \frac{1}{d})$

- $\mathbb{E}[\max(U, 1-U)] = \int_0^{1/2} 1-u du + \int_{1/2}^1 u du = 3/4$

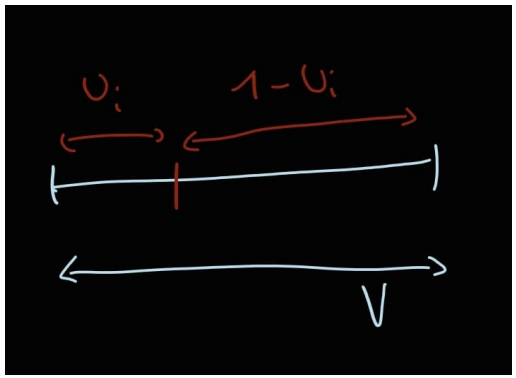


Figure 7 – Majoration d'une coupe par le plus grand des 2 segments.

Comme $K_n \sim \mathcal{B}(T_n, 1/d)$

$$
\begin{aligned}
\mathbb{E}\left[(3/4)^{K_n}\right] &= \mathbb{E}\left[\mathbb{E}\left[(3/4)^{K_n}\middle| T_n\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=0}^{t}(3/4)^i\,\mathbb{P}\left(K_n = i\right)\middle| T_n = t\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=0}^{t}(3/4)^i\binom{t}{i}(1/d)^i(1-1/d)^{t-i}\middle| T_n = t\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=0}^{t}(3/4d)^i\binom{t}{i}(1-1/d)^{t-i}\middle| T_n = t\right]\right] \\
&= \sum_{t\in T_n(\Omega)}\sum_{i=0}^{t}(3/4d)^i\binom{t}{i}(1-1/d)^{t-i}\,\mathbb{P}\left(T_n = t\right) \\
&= \sum_{t\in T_n(\Omega)}(1-1/d+3/4d)^t\,\mathbb{P}\left(T_n = t\right) \\
&= \mathbb{E}\left[(1-1/d+3/4d)^{T_n}\right]
\end{aligned}
$$

Thus, it suffices to show that $T_n \to \infty$ in probability. To this end, note that the partition tree is statistically related to a random binary search tree with $k+1$ external nodes (and thus k internal nodes). Such a tree is obtained as follows. Initially, the root is the sole external node, and there are no internal nodes. Select an external node uniformly at random, make it an internal node and give it two children, both external. Repeat until we have precisely k internal nodes and $k+1$ external nodes. The resulting tree is the random binary search tree on k internal nodes (see Devroye 1988 and Mahmoud 1992 for more equivalent constructions of random binary search trees). It is known that all levels up to $\ell = \lfloor 0.37 \log k \rfloor$ are full with probability tending to one as $k \to \infty$ (Devroye, 1986). The last full level F_n is called the fill-up level. Clearly, the partition tree has this property. Therefore, we know that all final cells have been cut at least ℓ times and therefore $T_n \geq \ell$ with probability converging to 1. This concludes the proof of Theorem 3.1. $\square$

- Identification à un arbre de recherche binaire (cf schéma ci-après) pour lequel les profondeurs $p \leq l \hat{=} \lfloor 0.37 \log k \rfloor$ sont totalement remplies avec une probabilité $\xrightarrow{k\infty} 1$ (Théorème).
- donc les feuilles (sur des niveaux pleins) sont donc de profondeur asymptotiquement $\geq l$ p.s. et donc asymptotiquement coupées au moins l fois.
- remarque on a la consistence malgré l'existence de cellules dont la dimension n°1 ne tend pas vers 0.

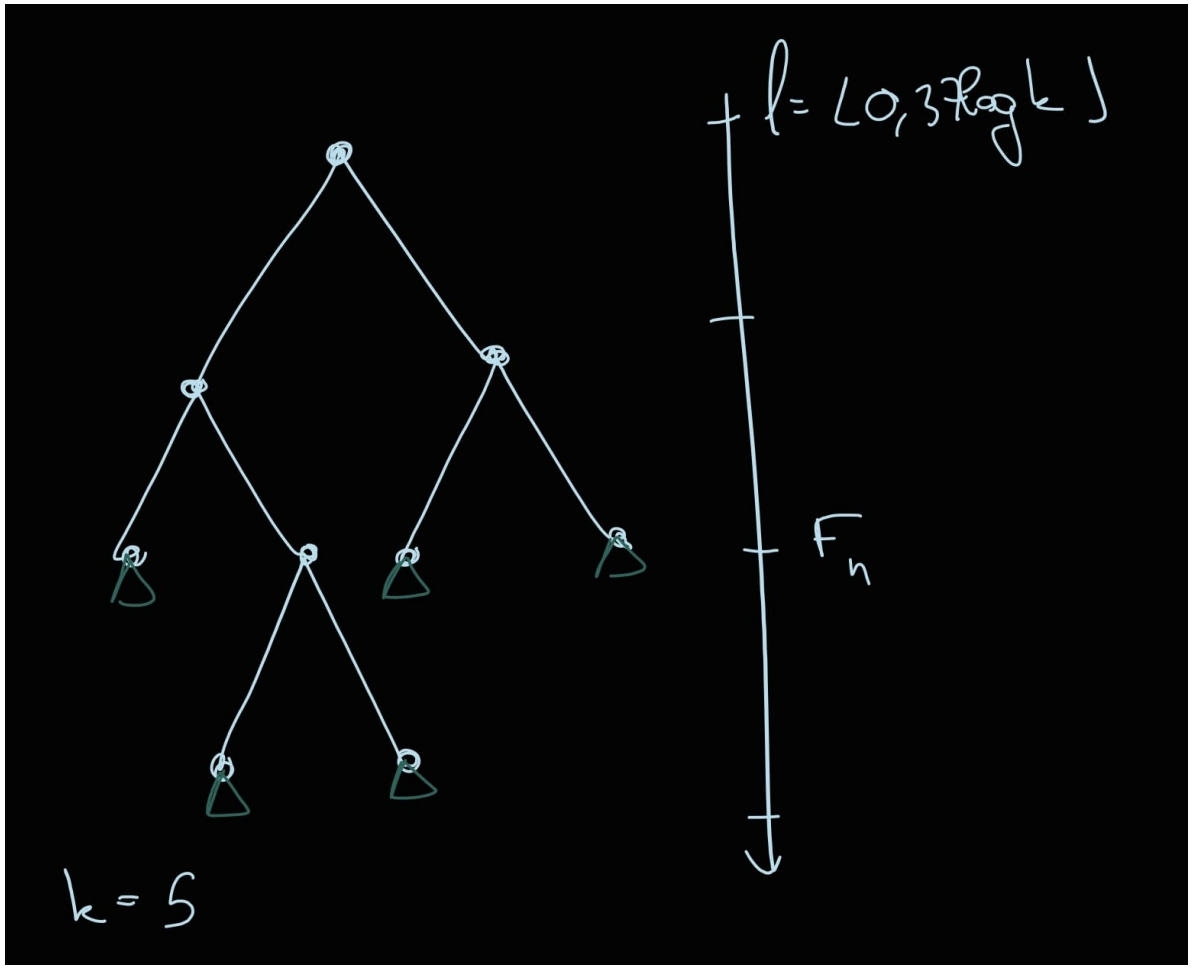


Figure 8 – Arbre binaire de recherche. On se souvient seulement des liens de "parenté" entre noeuds.

Forêt aléatoire invariante par échelle

Hypothèses :

- support de X dans $[0,1]^d \rightarrow$ Les $X^{(i)}$ ne sont pas atomiques, ie $\forall i, \mathbb{P}[X^{(i)} \in \{x\}] = 0$.

Construction :

- Cellule à couper C et direction de coupe $x^{(j)}$ choisies uniformément aléatoirement.
- Différence : abscisse de coupe faite en fonction du nombre d'échantillons $I \sim \mathcal{U}(\{0,1,...,n\})$ à placer dans l'une des 2 cellules enfants.

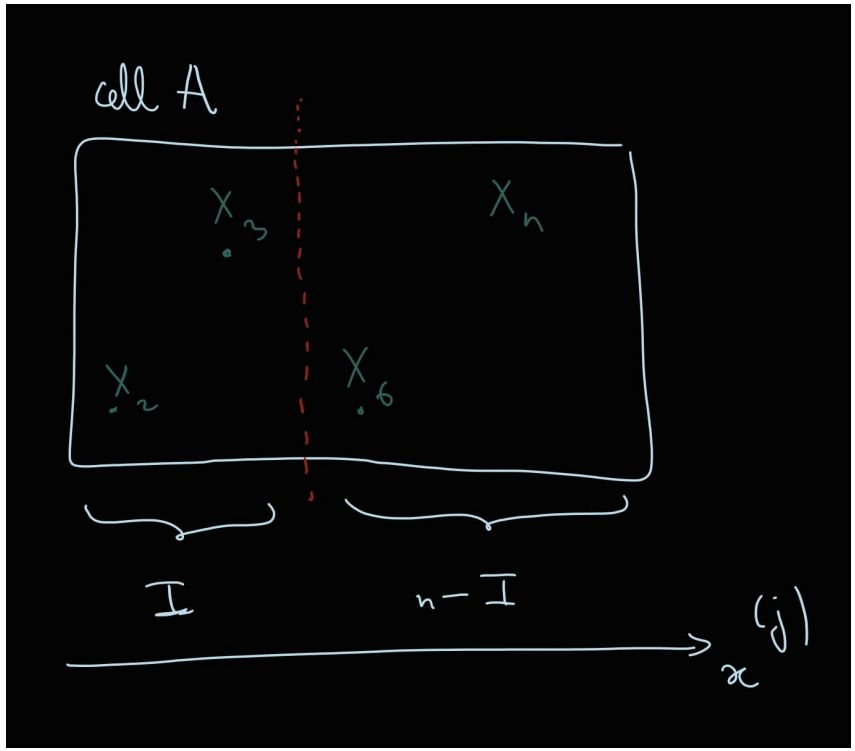


Figure 9 – Séparation d'une cellule selon un nombre I d'échantillons.

Théorème : Consistance des forêts invariantes par échelle

Si les $X^{(i)}$ ne sont pas atomiques, alors la forêt invariante par échelle $\overline{g_n}$ est consistante si $k_n \to \infty$ et $k_n/n \to \infty$ quand $n \to \infty$.

Preuve : Consistance des forêts invariantes par échelle

- $N_n(X, Z) \to \infty$ en proba : comme avant. (support de X n'avait pas été utilisé).

To show that $\text{diam}(A_n(X,Z)) \to 0$ in probability, we begin by noting that, just as in the case of the purely random forest classifier, the partition tree is equivalent to a binary search tree, and therefore with probability converging to one, all final cells have been cut at least $\ell = \lfloor 0.37 \log k \rfloor$ times.

Since the classification rule is scale-invariant, we may assume, without loss of generality, that the distribution of X is concentrated on the unit cube $[0,1]^d$.

Let n_i denote the cardinality of the i-th cell in the partition, $1 \le i \le k+1$, where the cardinality of a cell C is $|C \cap \{X, X_1, \ldots, X_n\}|$. Thus, $\sum_{i=1}^{k+1} n_i = n+1$. Let V_i be the first dimension of the i-th cell. Let $V(X)$ be the first dimension of the cell that contains X. Clearly, given the n_i's, $V(X) = V_i$ with probability $n_i/(n+1)$. We need to show that $\mathbb{E}[V(X)] \to 0$. But we have

$$\mathbb{E}[V(X)] = \mathbb{E}\left[\frac{\sum_{i=1}^{k+1} n_i V_i}{n+1}\right].$$

- Somme sur les i se fait parmi les noeuds externes. A relier à n le nombre total d'échantillons.

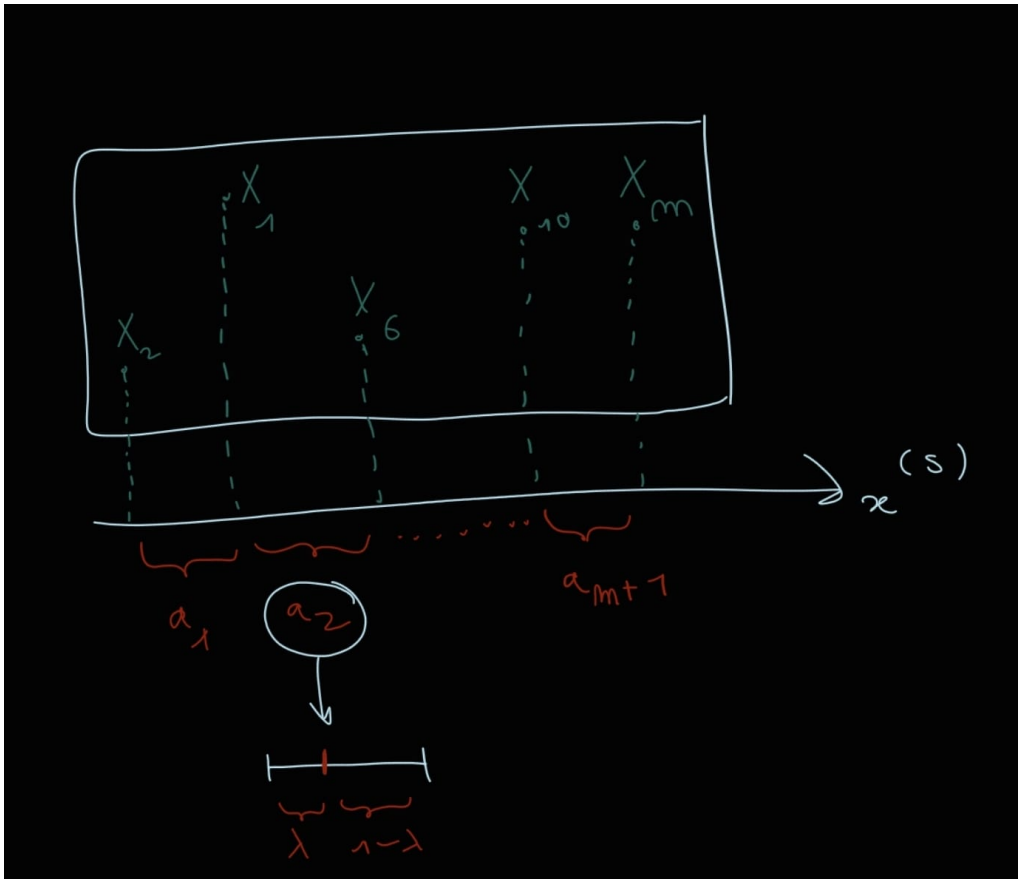


Figure 10 – Découpe d'une cellule par choix uniforme d'un espacement a_i et d'une proportion λ.

Une inégalité utile :
Si $A = \{$noeuds externes
du sous arbre de la cellule $i\}$ alors :

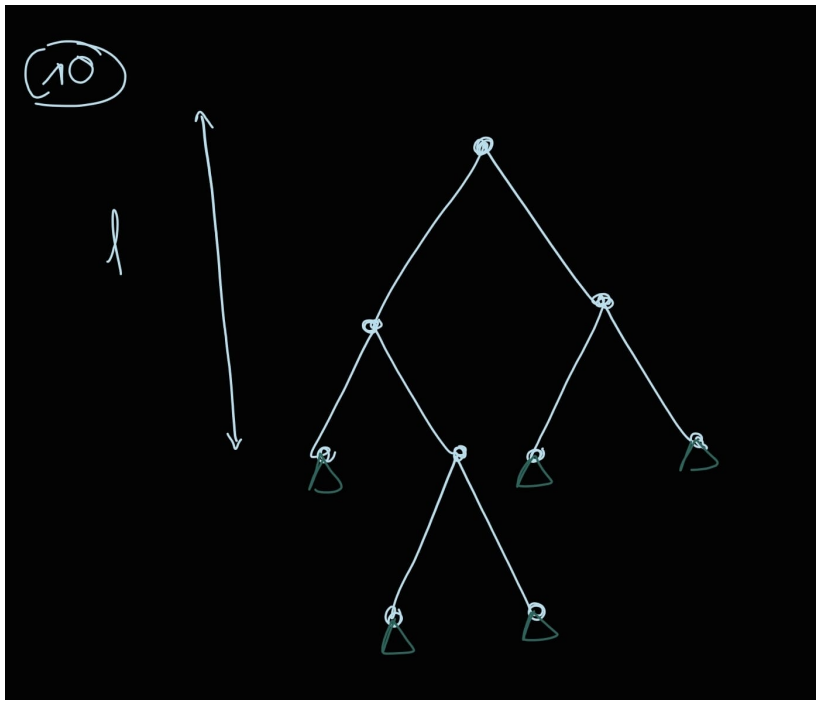
$$\begin{aligned}\sum_{j\in A} n_j V_j &=\leq \sqrt{\sum_j n_j^2}\sqrt{\sum_j V_j^2} \\ &\leq \sum_j n_j \sum_j V_j = n_i V_i \\ &\leq n\end{aligned}$$

Thus, if E is the collection of all external nodes of a partition tree, ℓ is at most the minimum path distance from any cell in E to the root, and L is the collection of all nodes at distance ℓ from the root, then, by the last inequality,

$$\sum_{i \in E} n_i V_i \leq \sum_{i \in L} n_i V_i.$$

Thus, using the notion of fill-up level F_n of the binary search tree, and setting $\ell = \lfloor 0.37 \log k \rfloor$, we have

$$\mathbb{E}\left[\sum_{i \in E} n_i V_i\right] \leq n\mathbb{P}\{F_n < \ell\} + \mathbb{E}\left[\sum_{i \in L} n_i V_i\right].$$



En effet,

Si $F_n > l \rightarrow$, alors E est plus profond que L et on partitionne les noeuds externes selon leur "ancêtre" dans L.

$$\mathbb{E}\left[\sum_{E} n_i V_i | F_n > l\right] \mathbb{P}(F_n \geq l)$$

$$\leq \mathbb{E}\left[\sum_{L} n_i V_i\right] \cdot 1$$

We have seen that the first term is $o(n)$. We argue that the second term is not more than $n(1 - 1/(8d))^{\ell}$, which is $o(n)$ since $k \to \infty$. That will conclude the proof.

It suffices now to argue recursively and fix one cell of cardinality n and first dimension V. Let C be the collection of its children. We will show that

$$\mathbb{E}\left[\sum_{i \in C} n_i V_i\right] \leq \left(1 - \frac{1}{8d}\right) nV.$$

Repeating this recursively ℓ times shows that

$$\mathbb{E}\left[\sum_{i \in L} n_i V_i\right] \leq n\left(1 - \frac{1}{8d}\right)^{\ell}$$

En effet

$$\begin{aligned}
\mathbb{E}\left[\sum_{i \in L} n_i V_i\right] &= \mathbb{E}\left[\sum_{j \in L-1} \sum_{i \text{ enfant de } j} n_i V_i\right] \\
&= \sum_{j \in L-1} \mathbb{E}\left[\sum_{i \text{ enfant de } j} n_i V_i\right] \\
&\leq \sum_{j \in L-1} (1 - (1/8d)) n_j V_j
\end{aligned}$$

Fix that cell of cardinality n, and assume without loss of generality that $V = 1$. Let the spacings along the first coordinate be $a_1, \ldots, a_{n+1}$, their sum being one. With probability $1 - 1/d$, there the first axis is not cut, and thus, $\sum_{i \in C} n_i V_i = n$. With probability $1/d$, the first axis is cut in two parts. We will show that conditional on the event that the first direction is cut,

$$\mathbb{E}\left[\sum_i n_i V_i\right] \leq \frac{7n}{8} .$$

Unconditionally, we have

$$\mathbb{E}\left[\sum_i n_i V_i\right] \leq \left(1 - \frac{1}{d}\right) n + \frac{1}{d} \cdot \frac{7n}{8} = \left(1 - \frac{1}{8d}\right) n ,$$

as required. So, let us prove the conditional result.

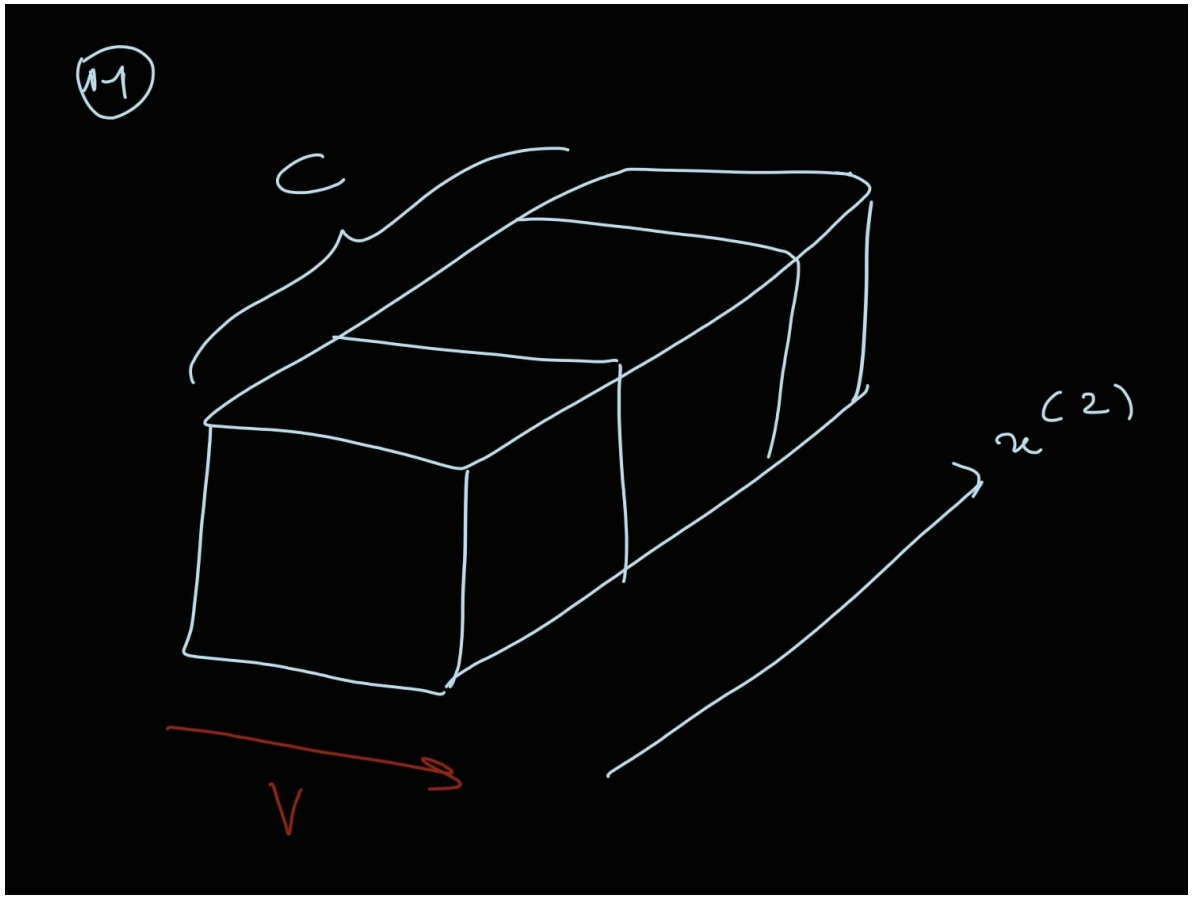


Figure 11 – Cellules filles découpées selon axe n°2

$$
\begin{aligned}
&\mathbb{E}\left[\sum_i n_i V_i\right] \\
&= \frac{1}{n+1}\mathbb{E}\Big[\sum_{j=1}^{n+1}\big[(j-1)(a_1+\cdots+a_{j-1}+a_j\delta_j) \\
&\qquad +(n+1-j)(a_j(1-\delta_j)+a_{j+1}+\cdots+a_{n+1})\big]\Big] \\
&= \frac{1}{n+1}\mathbb{E}\left[\sum_{k=1}^{n+1} a_k\left(\sum_{k<j\le n+1}(j-1)\right.\right. \\
&\qquad \left.\left.+\sum_{1\le j<k}(n+1-j)+\delta_k(k-1)+(1-\delta_k)(n+1-k)\right)\right] \\
&\le \frac{1}{n+1}\left(\sum_{k=1}^{n+1} a_k\left(n(n+1)-\frac{k(k-1)}{2}\right.\right. \\
&\qquad \left.\left.-\frac{(n-k+1)(n-k+2)}{2}+\max(k-1,n+1-k)\right)\right)
\end{aligned}
$$

- 1ère à 2e ligne : conditionnement selon l'espacement a_j qui est coupé.
- 3e à 4e ligne : formule de somme.

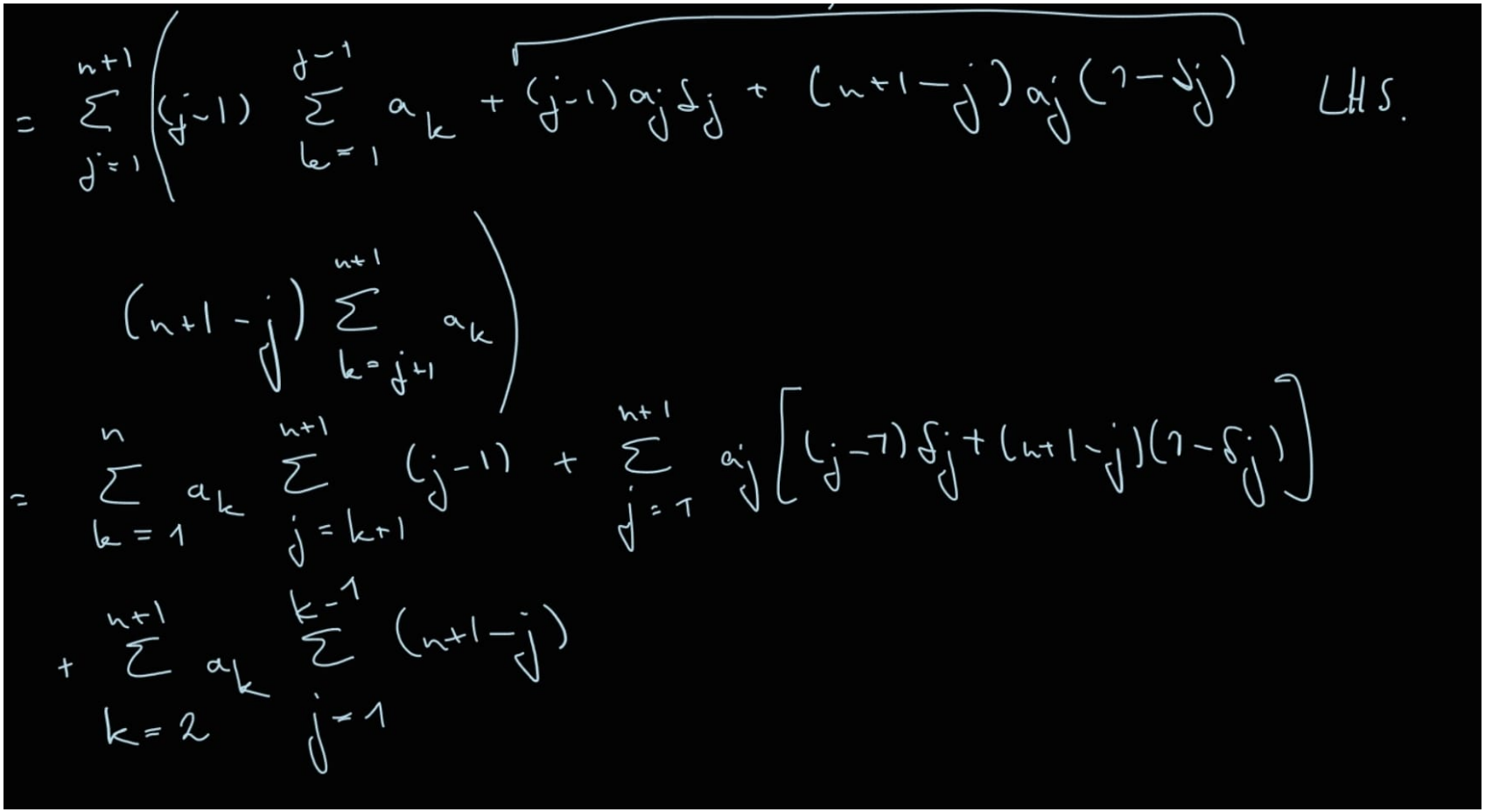


Figure 12 – 2e à 3e ligne : inversion de sommes.

$$
\begin{aligned}
&= \frac{1}{n+1}\left(\sum_{k=1}^{n+1} a_k \left(\frac{n(n+1)}{2} + (k-1)(n+1-k) + \max(k-1, n+1-k)\right)\right) \\
&\leq \frac{1}{n+1}\left(\left(\frac{n(n+1)}{2} + \left(\frac{n}{2}\right)^2 + n\right)\sum_{k=1}^{n+1} a_k\right) \\
&= n\left(\frac{3n/4 + (3/2)}{n+1}\right) \\
&\leq \frac{7n}{8} \quad \text{if } n > 4.
\end{aligned}
$$

□

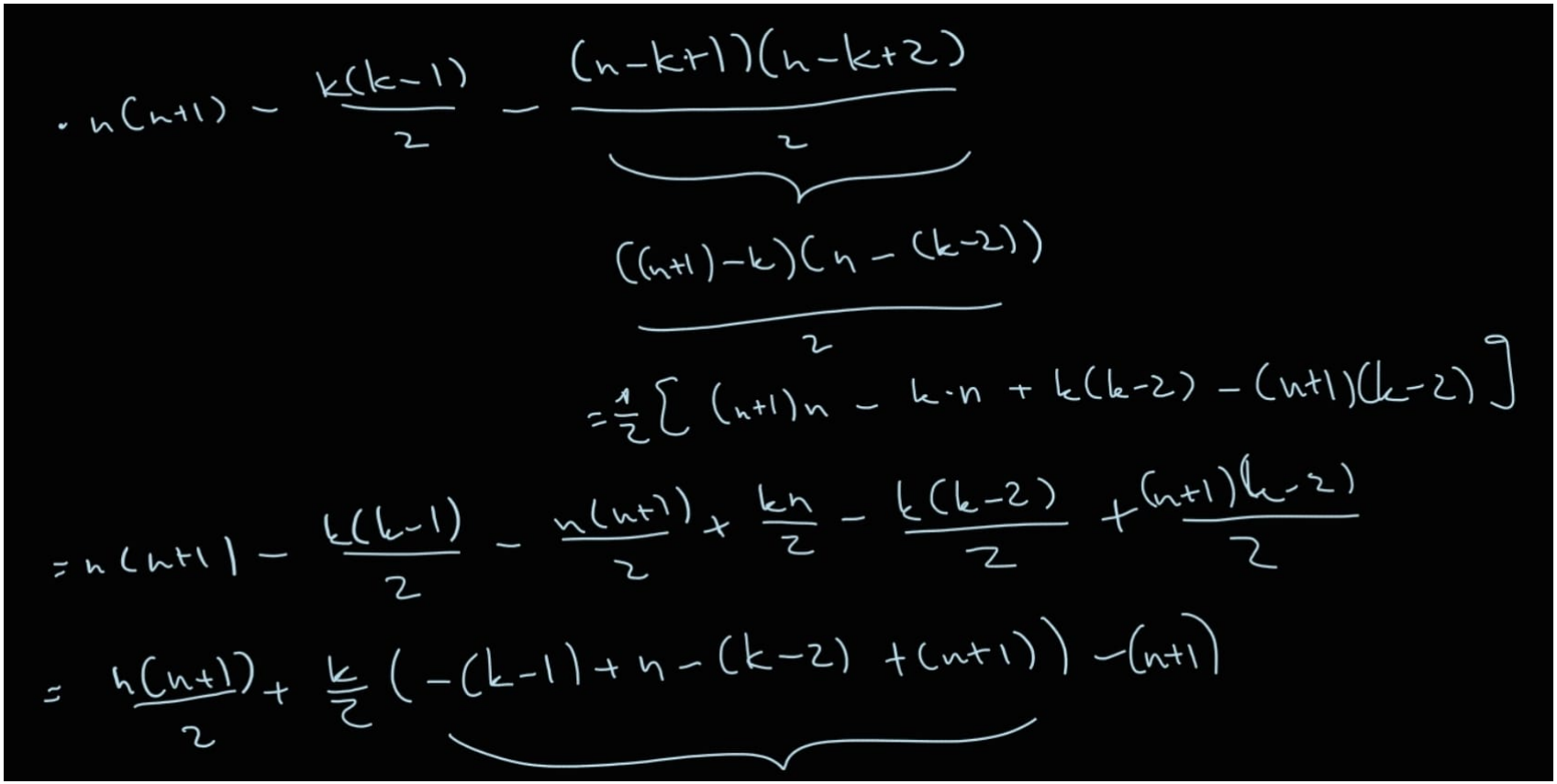


Figure 13 – Factorisation dans la dernière ligne (1)

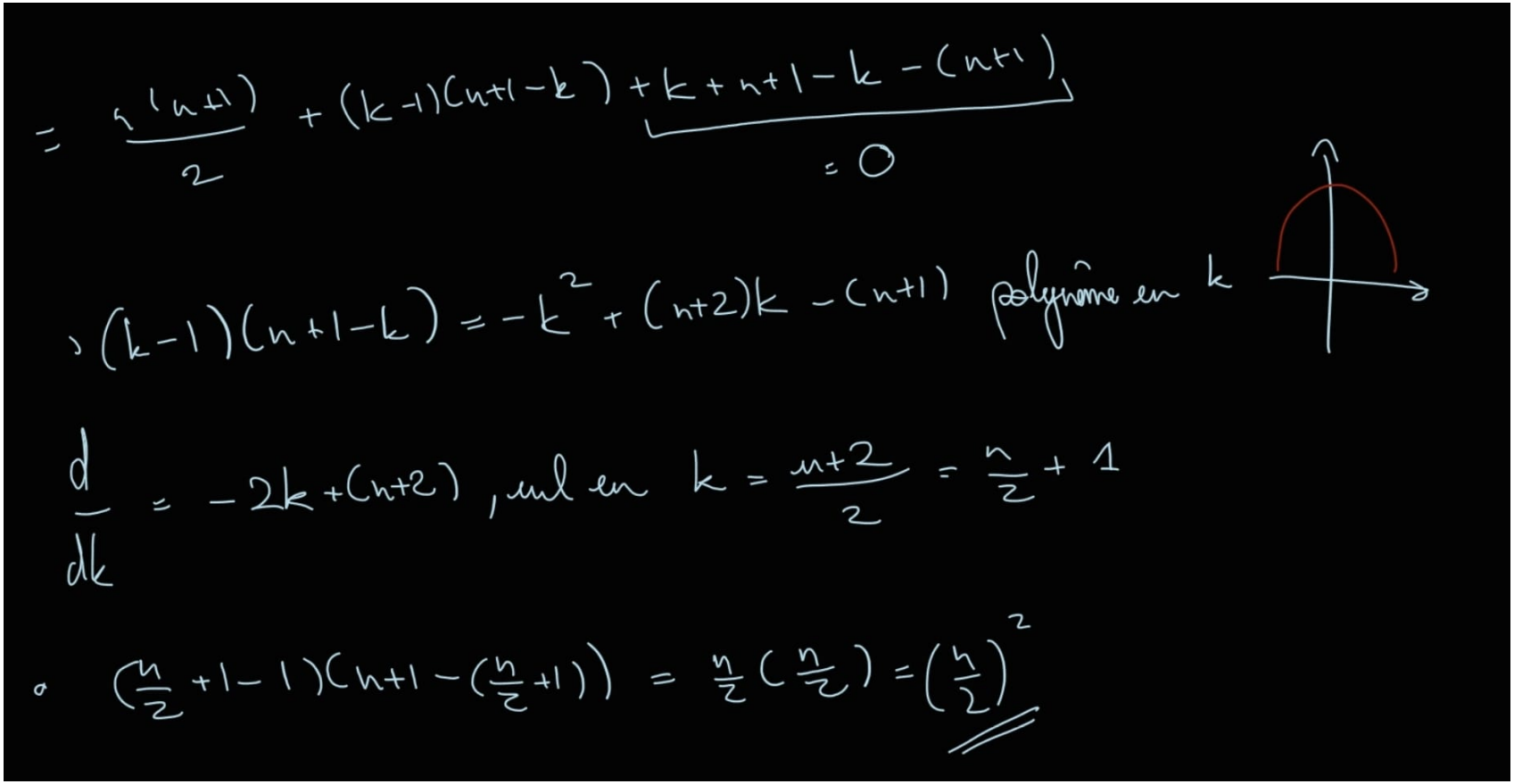


Figure 14 – Majoration par un polynôme en k

1 Introduction

2 Classifieurs de vote et de moyenne

3 Forêts Aléatoires

4 Consistence par Aggregation

5 Bootstrap Aggregation

6 Arbre à croissance gourmande

7 Conclusion

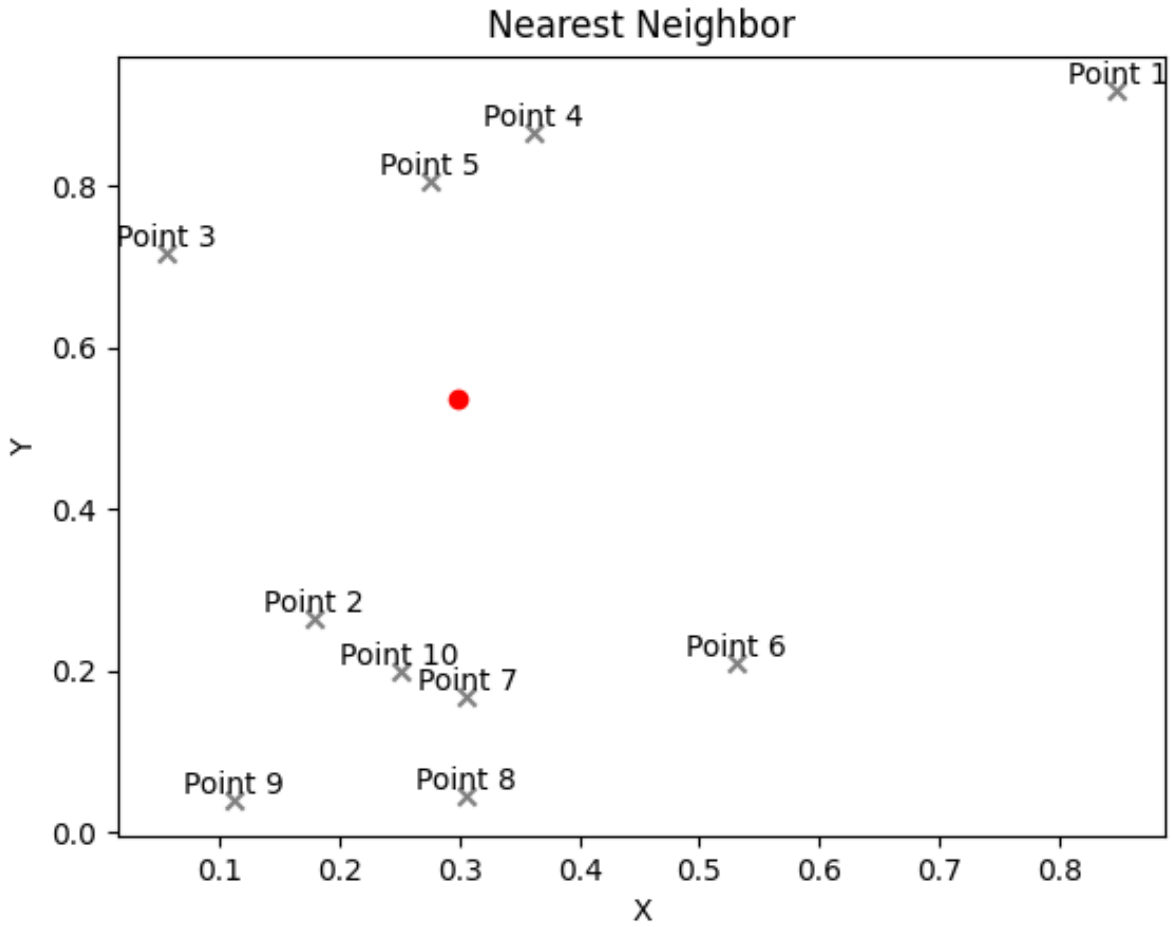
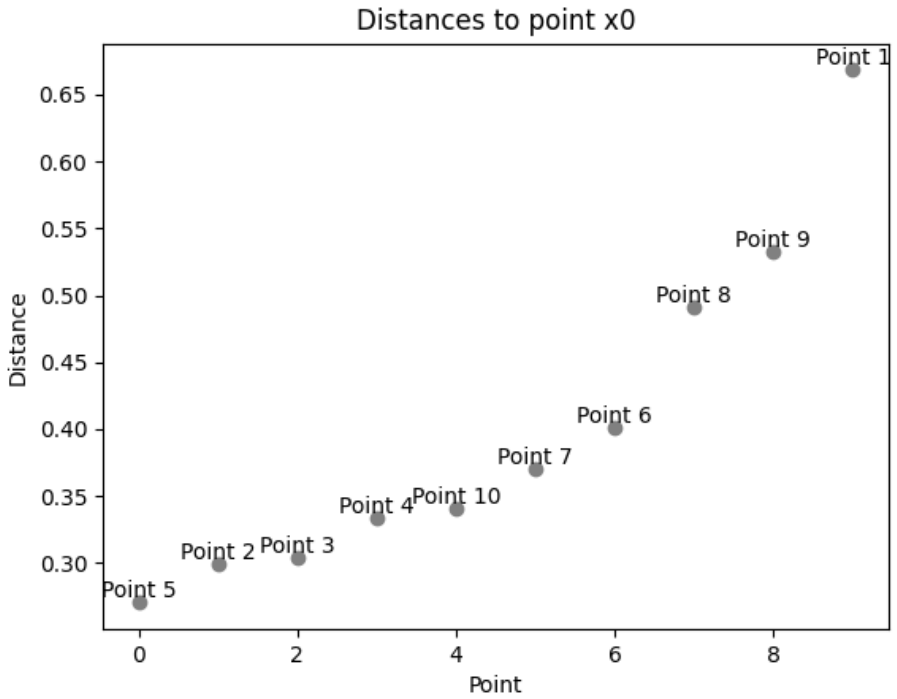


Figure 15 – Attribution du label



i,

$$\forall j \in [1, n], max(i, mU_i) \leq max(j, mU_j)$$

Théorème :

L'espérance du classifieur 1-plus proche voisin perturbé est consistant si $m \to \infty$ *et* $m/n \to 0$

1 Introduction

2 Classifieurs de vote et de moyenne

3 Forêts Aléatoires

4 Consistence par Aggregation

5 Bootstrap Aggregation

6 Arbre à croissance gourmande

7 Conclusion

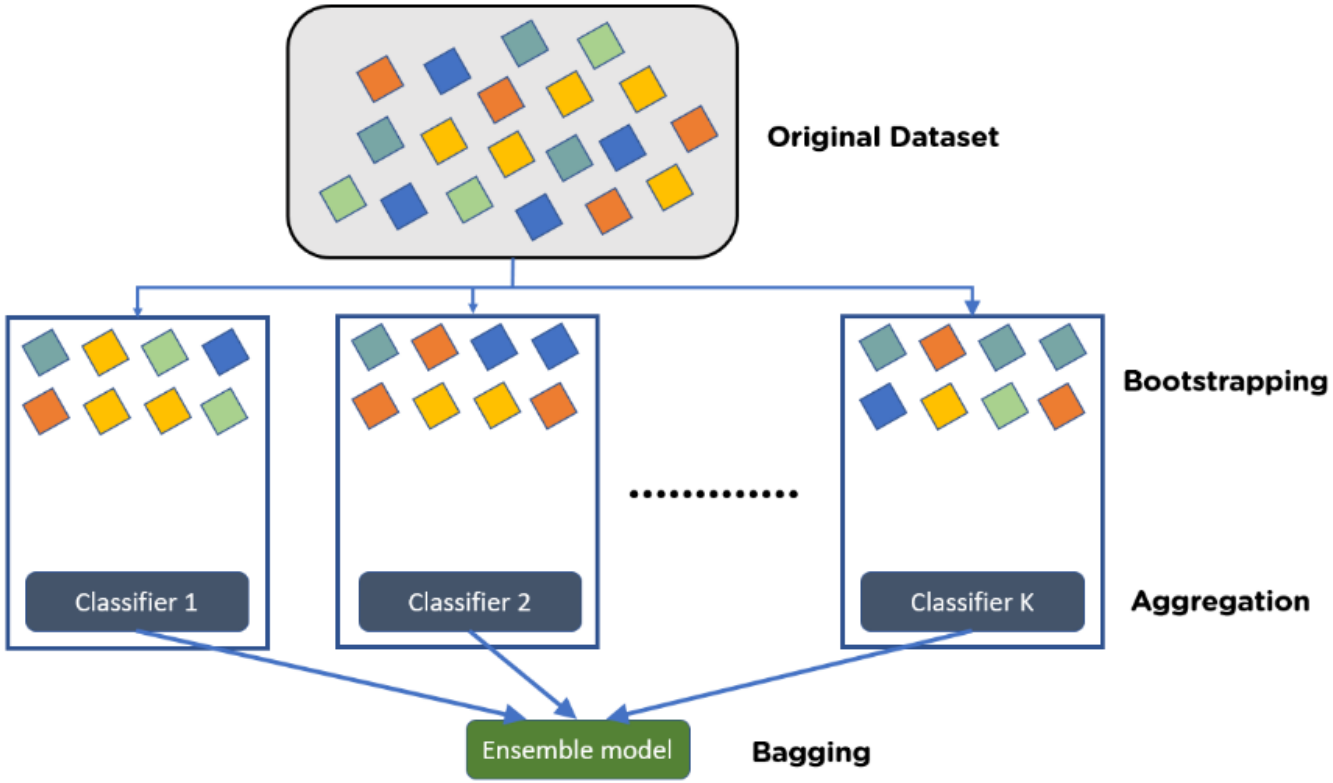


Figure 16 – Méthode de construction bagging

- $g_n(X, Z, D_n) = g_n(X, D_n(Z)) \rightarrow g_n(X, Z_m, D_n)$ (Classieur de vote)
- $g_n(x, D_n) = \mathbf{1}_{\{EZgN(x, D_n(Z)) \geq \frac{1}{2}\}}$ (Classifieur moyenné)
- Cas général : taille de l'échantillon bootstrap non constante et tirage sans remise
- Probabilité de présence d'un point donné : $q_n \in [0, 1]$

Théorème : Consistance d'un classificateur bagging

Si les classifieurs aléatoires g_n son consistant pour une distribution (X, Y) données, alors les classifieurs $g_n(x, Z_m, D_n)$ et $g_n(x, D_n)$ de paramètre q_n sont consistants si $nq_n \to \infty$ quand $n \to \infty$

- (Breiman) Tirage avec remise : $q_n \approx 1 - \frac{1}{e}$
- Des petites valeurs de qn peuvent transformer des classificateurs inconsistants en classificateurs consistants via le bagging.
- Resulat et illustration avec l'exemple du 1-plus proche voisin

Théorème : Consistance de la version bagging d'un kNN en fonction de q_n

Le classificateur du plus proche voisin $g_n(x, D_n)$ est consistant pour toutes les distributions (X,Y) si et seulement si $q_n \to 0 \quad$ et $\quad nq_n \to \infty$.

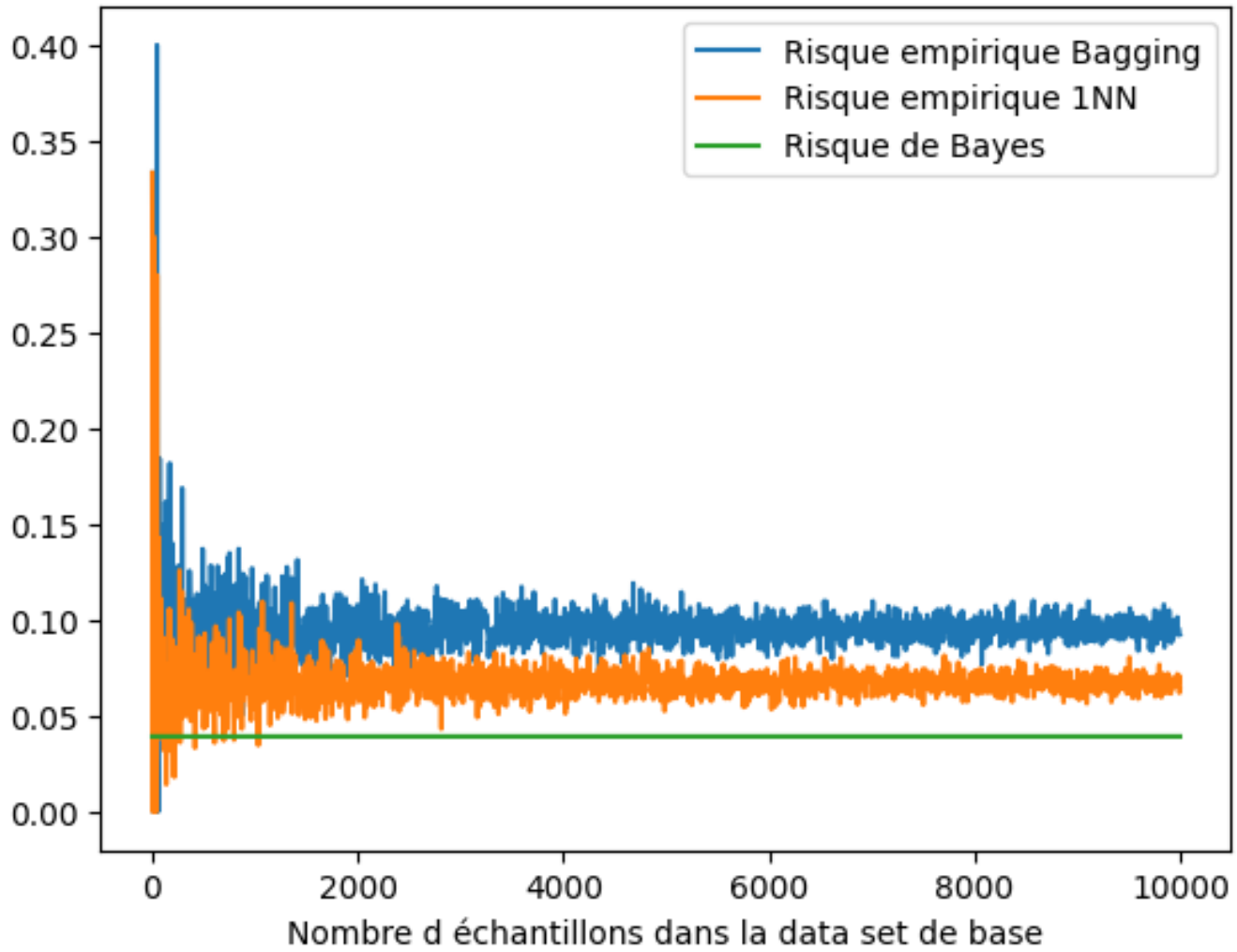


Figure 17 – Convergence classifieur 1-NN bagging en fonction de q_n

1 Introduction

2 Classifieurs de vote et de moyenne

3 Forêts Aléatoires

4 Consistence par Aggregation

5 Bootstrap Aggregation

6 Arbre à croissance gourmande

7 Conclusion

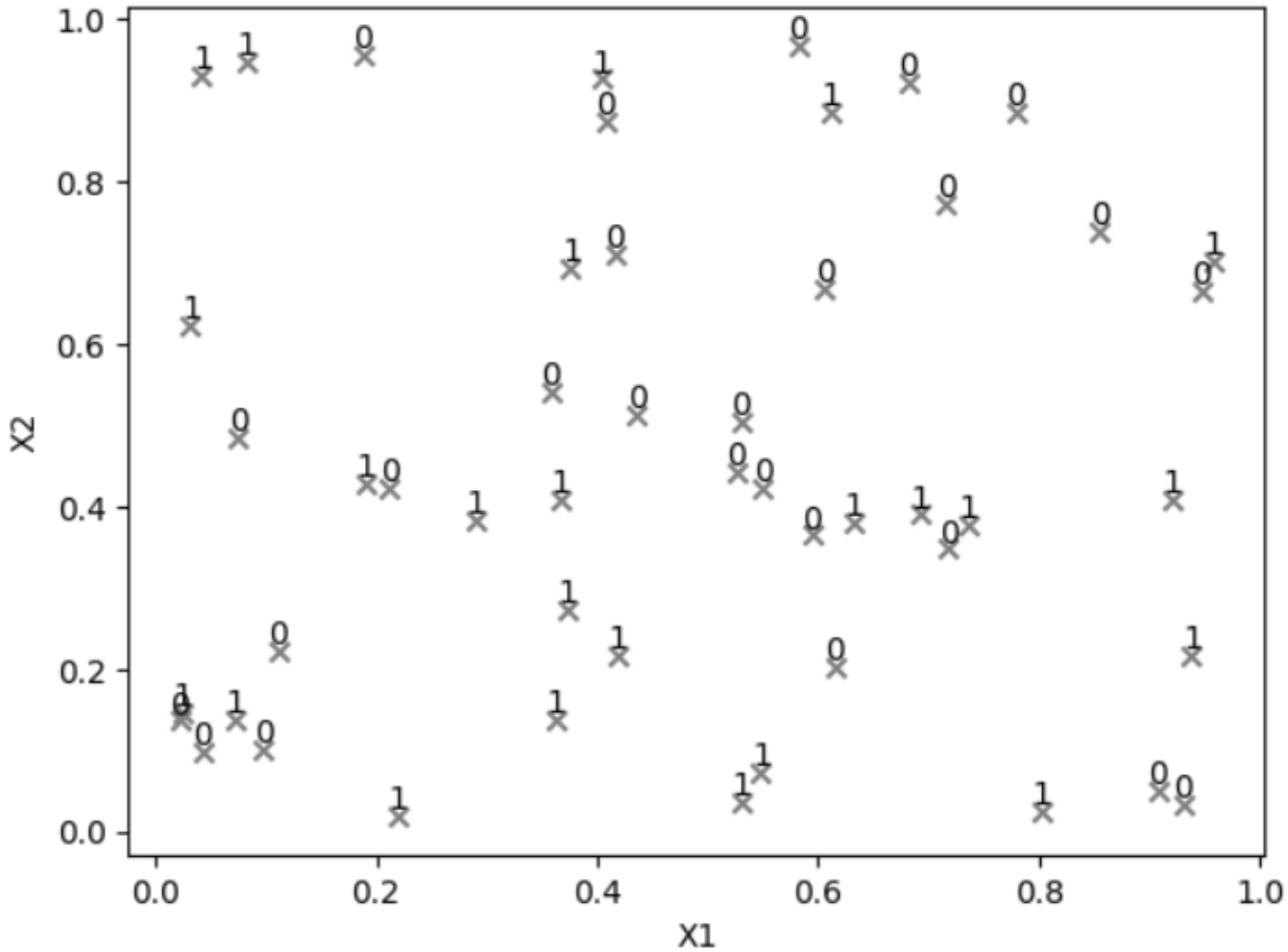


Figure 18 – Construction de l'arbre à croissance gourmande

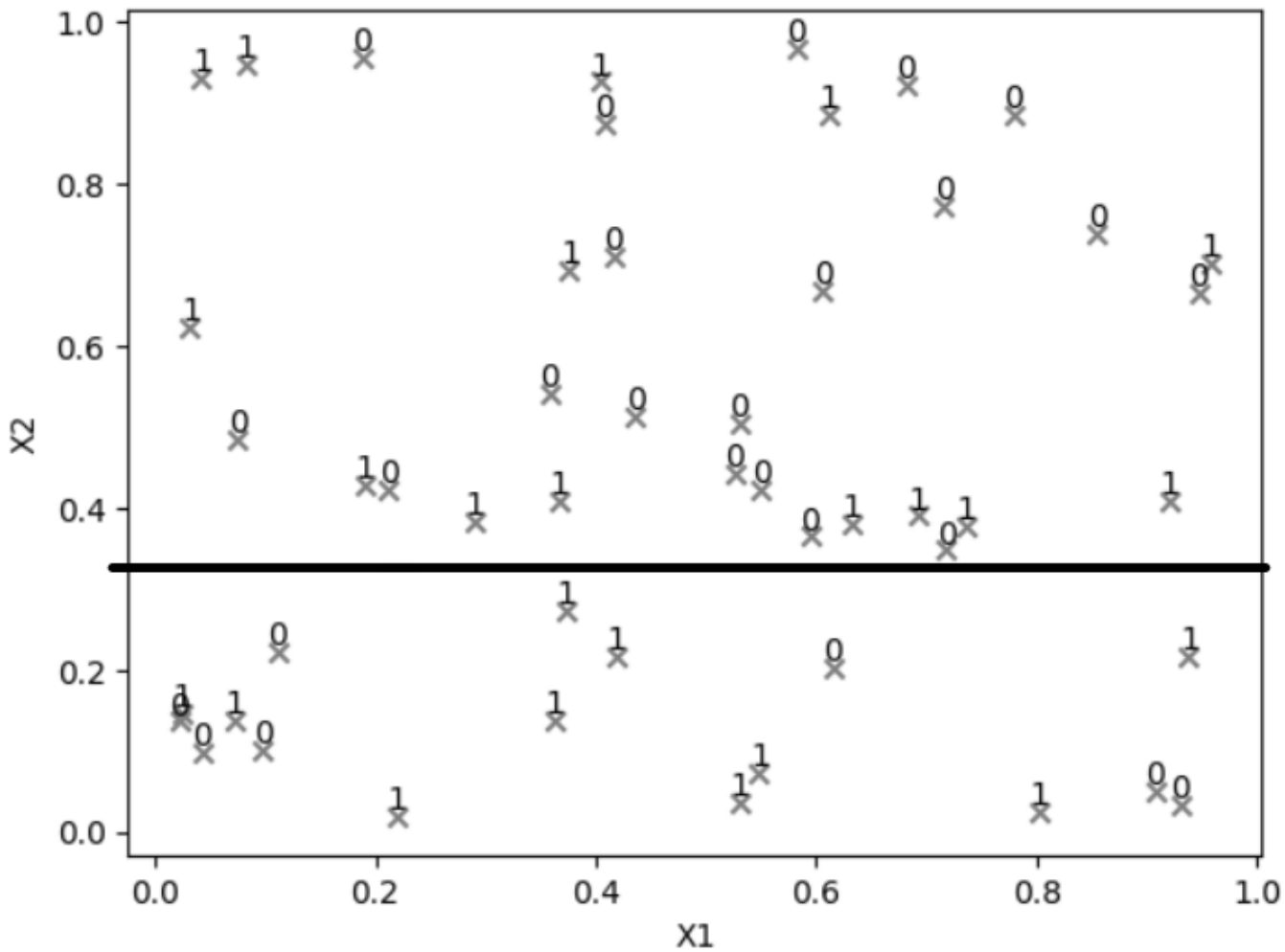


Figure 19 – Construction de l'arbre à croissance gourmande

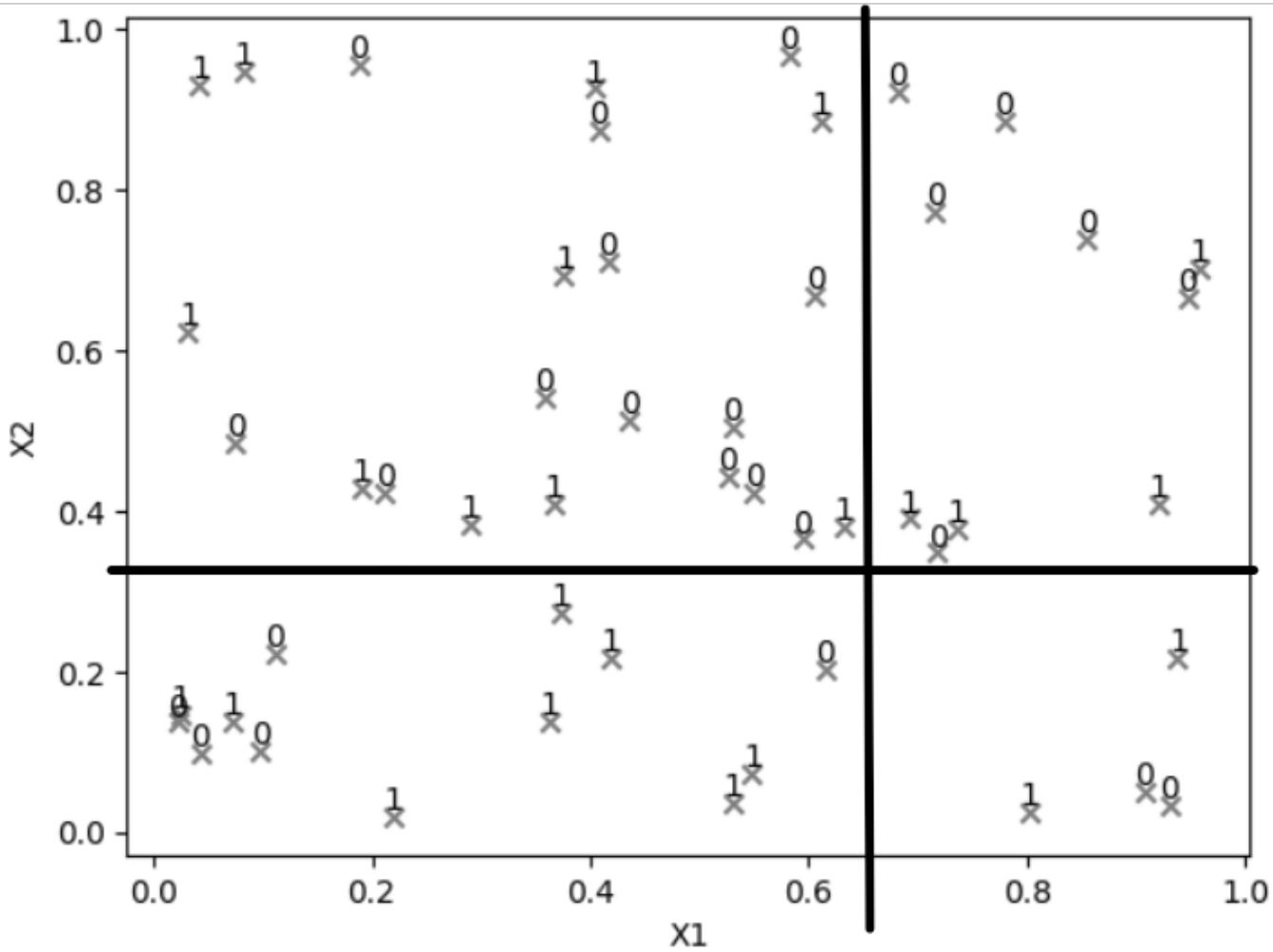


Figure 20 – Construction de l'arbre à croissance gourmande

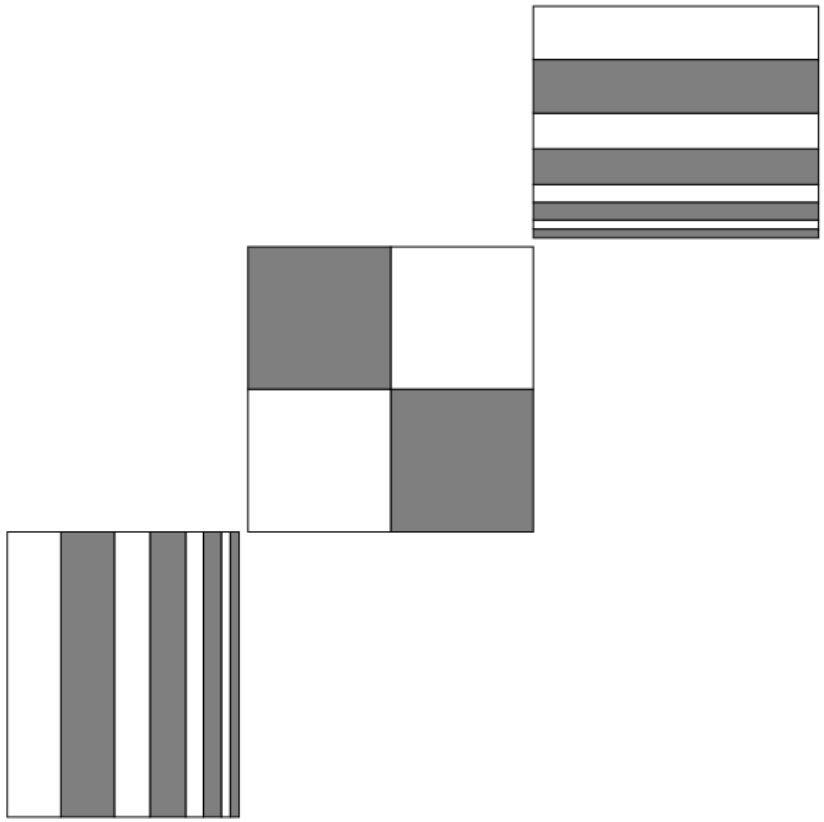


Figure 21 – Distribution pour laquelle l'arbre à croissance gourmande est inconsistant

Théorème :

Il existe des distributions pour lesquelles l'arbre à croissance gourmande est inconsistant

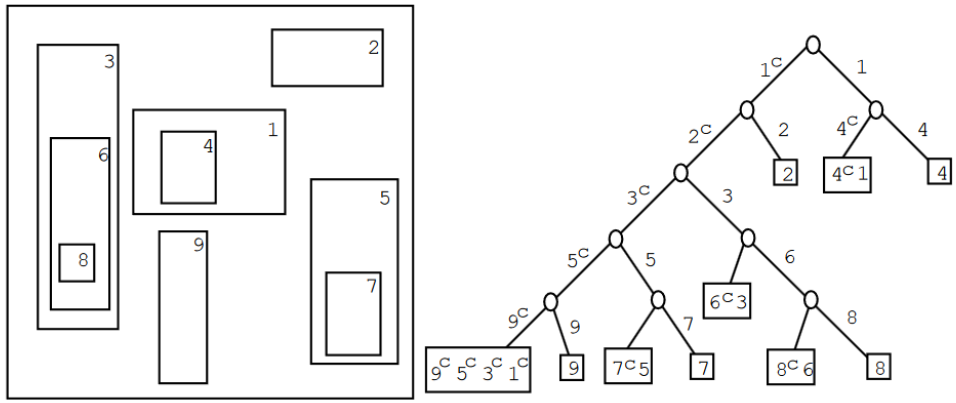


Figure 22 – Construction d'un algorithme gourmand consistant

On minimise à chaque étape $\hat{L}(T) + \hat{L}(R - T) - \hat{L}(R)$

Théorème :

L'arbre construit par hyper-rectangle optimisés est consistant avec $nq_n \to \infty, k \to \infty$ *et* $k = o(\sqrt{nq_n / ln(nq_n)})$

1 Introduction

2 Classifieurs de vote et de moyenne

3 Forêts Aléatoires

4 Consistence par Aggregation

5 Bootstrap Aggregation

6 Arbre à croissance gourmande

7 Conclusion

Conclusion : que cache la forêt ? Résultats importants

- Moyenner des classifieurs consistant conserve la consistance
- Moyenner des classifieurs inconsistant peut créer de la consistance
- Le bagging conserve la consistance des classifieurs
- Il peut servir à réduire les temps de calcul en entrainant de plus petits classifieurs sur moins de données