

MASTER MIAGE 2ÈME ANNÉE
UNIVERSITÉ PARIS NANTERRE

MÉMOIRE DE FIN D'ÉTUDES

Optimisation de parcours d'arbre de décisions dans le cadre du MNIST



Auteur :
ALEXANDRE PETIT-PAS

Tuteur :
P. MARIE-PIERRE
GERVAIS

Résumé

Résumé

Remerciements

Remerciements

Sommaire

1	Concepts formels	7
1.1	Introduction	7
1.2	Rappels de probabilité	7
1.3	Arbre de décision	8
1.4	Arbre de décision pour le classement	9
1.5	Arbre de décision pour l'optimisation	11
2	Lazy Decision Tree	13
2.1	Lazy Decision Tree pour le classement	13
2.2	Lazy Decision Tree pour l'optimisation	13
3	Classement supervisé	15
3.1	K Nearest Neighbors	15
3.2	CART	15
3.3	C4.5	15
3.4	A *	15
4	Apprentissage du MNIST	17
4.1	Définition du MNIST	17
4.2	Méthode de recherche	17
4.3	Algorithme de classement	17
4.4	Algorithme Lazy	17
4.5	Comparaison des résultats	17
	Bibliographie	19

Introduction

Motivations

Objectifs

Concepts formels

1.1 Introduction

L'objectif de ce chapitre est dans un premier temps de rappeler des notions de probabilités [1, 2, 3] qui serviront à la compréhension de ce document. Il s'agit ensuite d'expliquer dans sa généralité les concepts d'arbres décisionnels afin d'introduire des principes de classement et d'optimisation.

1.2 Rappels de probabilité

1.2.1 Expérience aléatoire

Une expérience est dite aléatoire si les résultats possibles sont connus à l'avance sans vraiment savoir celui obtenu au préalable. On appelle univers, noté Ω , l'ensemble de toutes les issues possibles d'une expérience et ω une réalisation de l'expérience.

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}, n \in \mathbb{N} \quad (1.1)$$

1.2.2 Événements

Un événement correspond à un ensemble de résultats possibles pour une expérience. Si cet ensemble est constitué d'un seul élément, on parle alors d'événement élémentaire. Si l'ensemble de résultats est égal à l'univers Ω , alors l'événement est dit certain. En revanche, si aucun résultat n'est présent (ensemble \emptyset), alors c'est un événement impossible.

1.2.3 Probabilités

Une probabilité est une fonction qui à un événement A , associe un poids.

$$\begin{cases} P(\Omega) = 1 \\ P(\emptyset) = 0 \\ 0 \leq P(A) \leq 1 \\ P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) \end{cases} \quad (1.2)$$

Plus la probabilité est proche de 1, plus il est possible que l'événement se réalise. Soit A et B deux événements quelconques, $P(A)$ est dite conditionnelle si son résultat est influencé par l'événement B :

$$\begin{cases} B \neq \emptyset \\ P(B) \neq 0 \\ A = (A \cap B) \cup (A \cap \overline{B}) \end{cases} \quad (1.3)$$

On peut alors déduire deux formules : le théorème de Bayes permettant de calculer $P_B(A)$ et le théorème des probabilités totales qui permet de connaître la valeur de $P(A)$ à partir de deux événements A et B.

$$P_B(A) = \frac{P(A \cap B)}{P(B)} \quad (1.4)$$

$$P(A) = P_B(A)P(B) + P_{\overline{B}}(A)P(\overline{B}) \quad (1.5)$$

1.3 Arbre de décision

1.3.1 Définition

Un arbre de décision est une représentation structurale qui permet d'aboutir à un choix. C'est un graphe acyclique orienté composé :

- d'un sommet sans parents appelé racine,
- de sommets appelés nœuds, correspondant à des tests,
- de sommets terminaux nommés feuilles,
- des arrêtes, ou branches, désignant chacune les résultats d'un test

Pour construire un arbre de décision, une approche *Top-Down* est utilisée, appelée *Top Down Induction of Decision Tree (TDIDT)*. Elle peut se décomposer en plusieurs parties :

1. Partir du jeu complet de données et construire la racine.
2. Réaliser un test afin de séparer les données.
3. Séparer le nœud actuel en fonction des résultats possibles.
4. Appliquer récursivement jusqu'à atteindre les feuilles.

Lorsque chaque nœuds sont composés exactement de deux descendants (hors feuilles), on parle alors d'arbre de décision binaire. C'est un cas très largement utilisé des algorithmes de construction d'arbres tels que CART [4], qui sera expliqué ultérieurement.

1.3.2 Exemple

Prenons comme exemple les données météorologiques [] présente dans la table 1. Cette table est constitué de différentes colonnes concernant différentes information ainsi qu'une colonne indiquant si la décision de sortir a été prise ou non. Dans un peu premier temps, il est possible de déduire les tests à réaliser comme par exemple

"La température est-elle élevée?" ou encore "Quel est le temps?". A ces questions découlent des réponses possibles, {oui, non} pour la première et {soleil, nuageux, pluie} pour la deuxième. Une fois l'approche TDIDT utilisée, un arbre de décision est alors obtenu (figure 1.1).

Jour	Temps	Température	Humidité	Vent	Sortie
1	soleil	élevée	haute	non	N
2	soleil	élevée	haute	oui	N
3	soleil	moyenne	haute	non	N
4	soleil	basse	normale	non	Y
5	soleil	moyenne	normale	oui	Y
6	nuageux	élevée	haute	non	Y
7	nuageux	basse	normale	oui	Y
8	nuageux	moyenne	haute	oui	Y
9	nuageux	élevée	normale	non	Y
10	pluie	moyenne	haute	non	Y
11	pluie	basse	normale	non	Y
12	pluie	basse	normale	oui	N
13	pluie	moyenne	normale	non	Y
14	pluie	moyenne	haute	oui	N

Table 1.1 – Table Météo

1.4 Arbre de décision pour le classement

1.4.1 Attributs et classes

Soit un ensemble d'observation S de taille n :

$$S = \{s_1, s_2, \dots, s_n\}.$$

Chaque observation s_i est composée d'un ensemble de m attributs :

$$A_s = \{a_1, a_2, \dots, a_m\}.$$

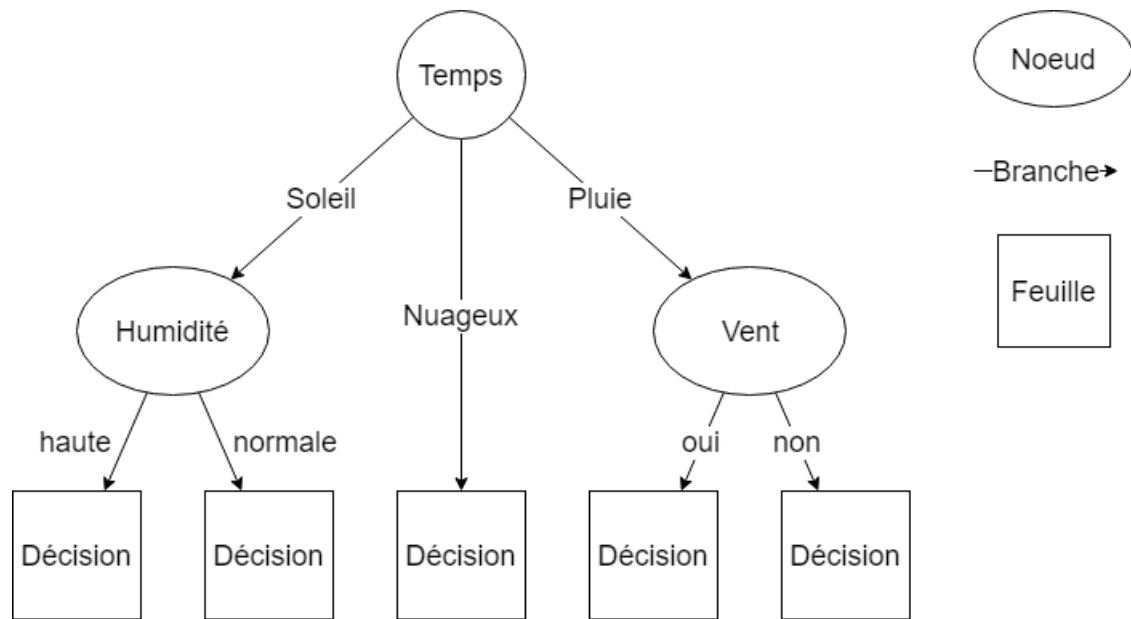


Figure 1.1 – Arbre de décision de la table Météo

Soit V_j l'ensemble de taille l des valeurs possibles de l'attribut a_j d'une observation s_i tels que :

$$V_j = \{v_1, v_2, \dots, v_l\}.$$

La valeur de l'attribut a_j sera représentée par v_k . Un attribut est dit qualitatif si l'ensemble des valeurs possibles est symbolique (non numérique), par exemple si V_j représente les couleurs d'écriture d'un mot. On obtiendrait alors $V_j = \{ \text{bleu, rouge, noir, vert} \}$.

Une donnée est quantitative si l'ensemble des valeurs possibles est un ensemble numérique fini ou infini :

- Si un attribut peut prendre une infinité de valeurs dans son ensemble, alors celui-ci est qualifié de continu, par exemple le temps d'exécution d'un processus.
- Dans le cas contraire, une variable dite discrète possède une valeur finie. Elle est généralement liée à une énumération, comme par exemple le nombre de trait dans un caractère.

a_j est nominal si la notion d'ordre n'est pas présente dans l'ensemble des valeurs possibles, par exemple si $V_j = \{ \text{un, deux, trois, quatre, cinq, six, sept, huit, neuf} \}$ représente le nom d'un chiffre en toutes lettres.

Un attribut est ordinal si la notion des valeurs possibles contiennent la notion d'ordre. Cela peut être par exemple l'appréciation d'un client : $V_j = \{ \text{mauvais, bon, très bon} \}$

Une variable est qualifiée de binaire si l'ensemble V_j des valeurs possibles est de taille $l = 2$

1.4.2 Fonction de classement

1.4.3 Gain d'information

1.4.4 Entropie

1.4.5 Information mutuelle

1.4.6 Exemple

1.5 Arbre de décision pour l'optimisation

Lazy Decision Tree

2.1 Lazy Decision Tree pour le classement

2.2 Lazy Decision Tree pour l'optimisation

Classement supervisé

3.1 K Nearest Neighbors

3.2 CART

3.3 C4.5

3.4 A *

Apprentissage du MNIST

- 4.1** Définition du MNIST
- 4.2** Méthode de recherche
- 4.3** Algorithme de classement
- 4.4** Algorithme Lazy
- 4.5** Comparaison des résultats

Bibliographie

- [1] Hélène Guérin. *Qu'est ce qu'une probabilité ?* 2008.
- [2] Lamis Hawarah. "Une approche probabiliste pour le classement d'objets incomplètement connus dans un arbre de décision". Thèse de doct. Université Joseph Fourier - Grenoble I, 2008.
- [3] R. Gilleron F. Denis. *Apprentissage à partir d'exemples*. 2000.
- [4] Jan Kozak. *Decision tree and ensemble learning based on ant colony optimization*. Springer, 2019.

Table des matières

1	Concepts formels	7
1.1	Introduction	7
1.2	Rappels de probabilité	7
1.2.1	Expérience aléatoire	7
1.2.2	Événements	7
1.2.3	Probabilités	7
1.3	Arbre de décision	8
1.3.1	Définition	8
1.3.2	Exemple	8
1.4	Arbre de décision pour le classement	9
1.4.1	Attributs et classes	9
1.4.2	Fonction de classement	11
1.4.3	Gain d'information	11
1.4.4	Entropie	11
1.4.5	Information mutuelle	11
1.4.6	Exemple	11
1.5	Arbre de décision pour l'optimisation	11
2	Lazy Decision Tree	13
2.1	Lazy Decision Tree pour le classement	13
2.2	Lazy Decision Tree pour l'optimisation	13
3	Classement supervisé	15
3.1	K Nearest Neighbors	15
3.2	CART	15
3.3	C4.5	15
3.4	A *	15

4	Apprentissage du MNIST	17
4.1	Définition du MNIST	17
4.2	Méthode de recherche	17
4.3	Algorithme de classement	17
4.4	Algorithme Lazy	17
4.5	Comparaison des résultats	17
	Bibliographie	19

Table des figures

1.1	Arbre de décision de la table Météo	10
-----	---	----

Liste des tableaux

1.1 Table Météo 9