

Arbres de Décisions

L'apprentissage automatique (2016-2017)

UFR MIME

Université Lille 3

6 Mars 2017

Sommaire

Arbres de
Décisions

Lille 3

Introduction

Entropie et
gain
d'information

Structure et
construction
d'arbre

1 Introduction

2 Entropie et gain d'information

3 Structure et construction d'arbre

Akinator

Arbres de
Décisions

Lille 3

Introduction

Entropie et
gain
d'information

Structure et
construction
d'arbre



<http://fr.akinator.com/personnages/>

Exemple simple

Arbres de
Décisions

Lille 3

Introduction

Entropie et
gain
d'information

Structure et
construction
d'arbre

- Donnee qualitatives

Patient	Toux	Fievre	Poids	Douleur	Diagnostic
Marie	non	oui	normal	gorgle	rhume
Fred	non	oui	normal	abdomen	Appendicite

- Attributs : Toux, Fievre, Poids, Douleur, Diagnostic
- Classes a predire : rhume, Appendicite

Exemple simple

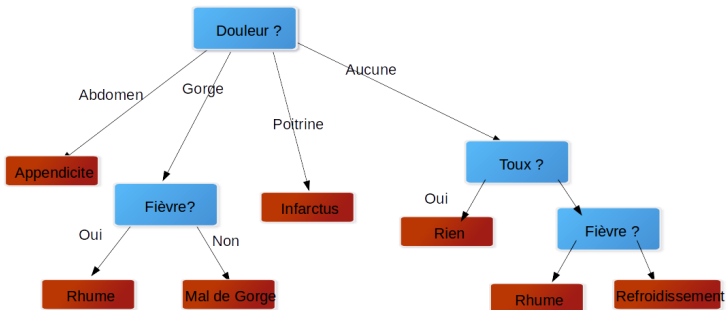
Arbres de
Décisions

Lille 3

Introduction

Entropie et
gain
d'information

Structure et
construction
d'arbre



- Nœuds Bleu : nœuds de decisions
- Nœuds Rouge : nœuds de predictions

Arbres de Decision

Arbres de
Décisions

Lille 3

Introduction

Entropie et
gain
d'information

Structure et
construction
d'arbre

- Un arbre de decision est un classifieur représenté sous forme d'un arbre.
 - Chaque nœuds testent les attributs pour determiner la branche
 - Il y a une chemin dans l'arbre associer à chaque nouvelle donnée (vecteur)
 - Quand on arrive a les feuilles, la prédiction par le classifieur est faites par une vote majoritaire parmi toutes les autres échantillonions qui sont achever cette feuille.
- Les arbres de décision constituent une méthode efficace d'exploration de données, en vue de la prédiction (classification et regression) d'une variable à l'aide de variables de tout type qualitatives et/ou quantitatives.

Calcul de la prédiction/classe

Arbres de Décisions

Lille 3

Introduction

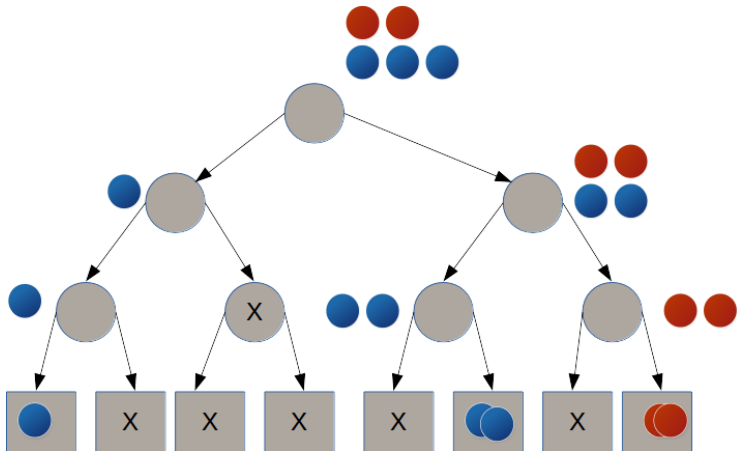


FIGURE : Arbre deja construite.

Creation d'arbre

Arbres de
Décisions

Lille 3

Introduction

Entropie et
gain
d'information

Structure et
construction
d'arbre

- Choisir un axe ou dimension
- **Comment choisir la bonne axe ?**
- Choisir une valeur sur l'axe a laquelle on va decouper l'espace.
- **pour que la partition resultant a des fils que contient des echantillon d'apprentissage de la meme classe.**
- **Comment mesurer la qualite de decoupage ?**
- Decouper l'espace en deux (pour arbre binaire) partie ou sous-ensemble.
- Repeter les premiers trois etapes sur les sous-ensemble obtenue. Continuer jusqu'a une critere de terminaison : le nœuds contiens que les points d'un classe.

Entropie

Arbres de
Décisions

Lille 3

Introduction

Entropie et
gain
d'information

Structure et
construction
d'arbre

- Évaluation de l'apport d'une découpage (décision base sur une attribut) par **gain d'information**

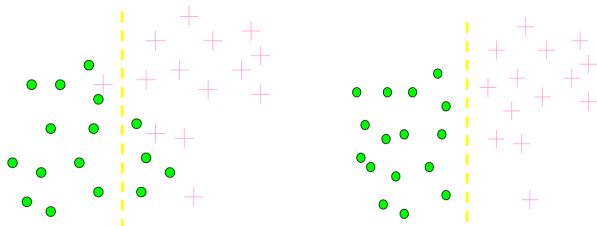


FIGURE : Choix entre deux decoupages possible : gauche, droite.

- Retrouver quelle decision vas **bien separer** les points.
- Pour quoi separer les points ?

Entropie

Arbres de
Décisions

Lille 3

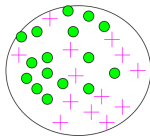
Introduction

Entropie et
gain
d'information

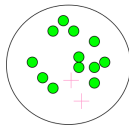
Structure et
construction
d'arbre

- Comment mesurer quelles sequences des question/decision ont les bonnes capacités de bien predires les classes.
- Comment determiner dans un noeud de decision, quelle attribut/coordonee parmi toutes les coordonees d'un vecteur est utile pour bien separer les classes.

Very impure group



Less impure



**Minimum
impurity**

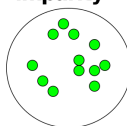


FIGURE : Different niveau de melanges obtenue par different choix de decoupage (axes et valeurs).

Gain d'information

Arbres de
Décisions

Lille 3

Introduction

Entropie et
gain
d'information

Structure et
construction
d'arbre

- Entropie d'un ensemble S est
$$H(S) = - \sum_{c \in \{o, +\}} p_c \log_2 p_c$$
- $\{o, +\}$ sont des classes ou valeur cible a predire pour la classificaion.
- Vous avez d axes parmi la quelle il faut choisir. Chaque axes est denommer x_j
- Etant donnee une ensemble S et un axe axe x_j on doit associer a chaque attribut a le Gain d'information :
- $G(S, a) = H(S) - \sum_{a \in \text{valeur unique}(A)} \frac{|S_a|}{|S|} H(S_a)$
- $G(S, a) = \text{Entropie}(\text{Parent}) - \text{Entropie Moyenne}(\text{Fils produit par attribut } a)$

Exemple

Arbres de
Décisions

Lille 3

Introduction

Entropie et
gain
d'information

Structure et
construction
d'arbre

X1	X2	X3	C
rouge	cercle	lourd	A
rouge	cercle	légère	A
bleu	carre	lourd	B
rouge	carre	légère	B

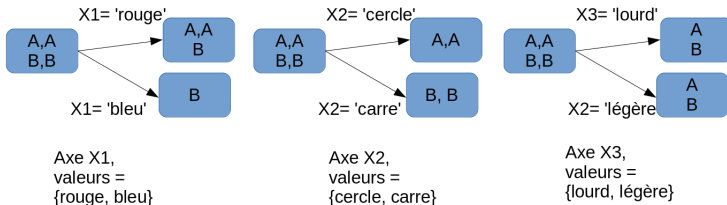


FIGURE : Quelle decoupage (axes+ valeur) est mieux ?

Algorithme (ID3)

Arbres de
Décisions

Lille 3

Introduction

Entropie et
gain
d'information

Structure et
construction
d'arbre

Entree : $(X_{\text{train}}, y_{\text{train}})$, variable categorique.

- Diviser/Splitter (noeud, exemples (X, y))
- A = la meilleure attribut/axe qui peut separer exemples avec gain d'information
- Fixer Axe = A pour cette noeud
- Pour chaque valuer unique a dans l'axe A dans le donnee d'entrainement, extraire les index ou les valeur de $X_{\text{train}}[?, A] = a$. Ces indcies sont les fils de noeud
- Pour chaque fils/sous-ensemble
 - si sous-ensemble est **pur** : STOP
 - Diviser/Splitter (fils, sous-ensemble)

Donnee pour aujourd'hui

Arbres de
Décisions

Lille 3

Introduction

Entropie et
gain
d'information

Structure et
construction
d'arbre

Attributs	Pif	Temp	Humid	Vent
Valeurs possibles	soleil,couvert,pluie	chaud,bon,frais	normale,haute	vrai,faux

N°	Pif	Temp	Humid	Vent	Golf
1	soleil	chaud	haute	faux	NePasJouer
2	soleil	chaud	haute	vrai	NePasJouer
3	couvert	chaud	haute	faux	Jouer
4	pluie	bon	haute	faux	Jouer
5	pluie	frais	normale	faux	Jouer
6	pluie	frais	normale	vrai	NePasJouer
7	couvert	frais	normale	vrai	Jouer
8	soleil	bon	haute	faux	NePasJouer
9	soleil	frais	normale	faux	Jouer
10	pluie	bon	normale	faux	Jouer
11	soleil	bon	normale	vrai	Jouer
12	couvert	bon	haute	vrai	Jouer
13	couvert	chaud	normale	faux	Jouer
14	pluie	bon	haute	vrai	NePasJouer

la classe

Arbre obtenu

Arbres de
Décisions

Lille 3

Introduction

Entropie et
gain
d'information

Structure et
construction
d'arbre

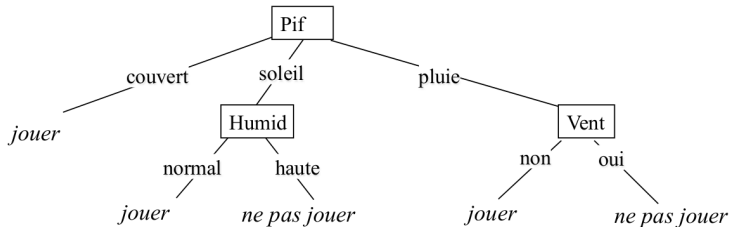


FIGURE : Solution

Comment predire maintenant pour un nouveau echantillons :
[Jour15, 'pluie', 'haut', 'faible']

Prediction/Classification

Arbres de
Décisions

Lille 3

Introduction

Entropie et
gain
d'information

Structure et
construction
d'arbre

Il nous reste encore :

- Etant donnee un nouveau echantillons retrouver le chemins dans l'arbre en comparent les.
- Etant donnee la feuille associer a un echantillon, predire le vote majoritaires des echantillons dans cette feuilles.
- Comment travailler avec les attributs/axes numerique et continue ?

