

IA et science des données

Cours 7 – mardi 8 mars 2022
Arbres de décision (suite)

Christophe Marsala
Vincent Guigue

Sorbonne Université

LU3IN026 - 2021-2022

Plan du cours

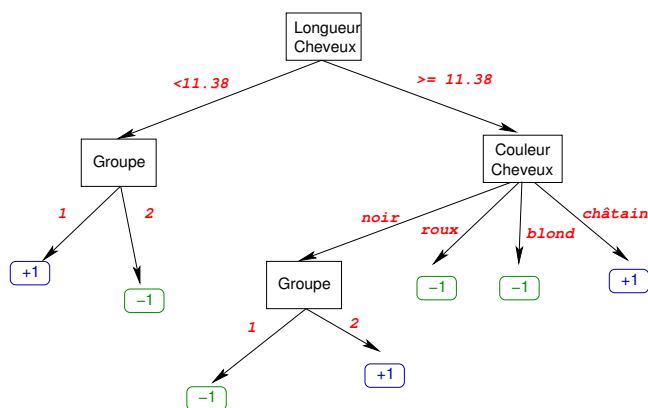
Apprentissage par arbres de décision (suite)

frontière
Classification et interprétabilité
Conclusion

Méthodes d'ensembles

1 – Apprentissage par arbres de décision (suite) –

Exemple d'arbre de décision



Marsala & Guigue – 2022

LU3IN026 – cours 7 – 3

1 – Apprentissage par arbres de décision (suite) –

Critère d'arrêt de la construction de l'arbre

- Quelques exemples de critères d'arrêt
 - tous les exemples de la base d'apprentissage ont la même classe
 - utilisation d'une tolérance : la **plupart** des exemples ont la même classe
 - utilisation d'un seuil $\varepsilon \in [0, 1]$
 - on calcule $H(\mathbf{Y}) = -\sum_k p(y_k) \log p(y_k)$ et on s'arrête si $H(\mathbf{Y}) \leq \varepsilon$
 - le gain d'information est nul ou négatif : $I_S(\mathbf{X}_j, \mathbf{Y}) \leq 0$
 - trop peu d'exemples dans l'ensemble traité
 - cas catégoriel : tous les attributs ont été utilisés une fois
- Création d'une **feuille** de l'arbre de décision
 - la classe majoritaire est utilisée pour étiqueter la feuille

Marsala & Guigue – 2022

LU3IN026 – cours 7 – 5

1 – Apprentissage par arbres de décision (suite) –

Construction de l'arbre : algorithme classique (catégoriel)

- Créer une pile \mathcal{P} et y stocker la base d'apprentissage
- Tant que \mathcal{P} n'est pas vide : prendre l'ensemble \mathcal{E} en haut de \mathcal{P}
 - calculer $H(\mathbf{Y})$ pour \mathcal{E}
 - si le critère d'arrêt est atteint alors créer une feuille
 - sinon, pour les exemples de \mathcal{E}
 - calculer $H(\mathbf{Y}|\mathbf{X}_j)$ pour tous les attributs \mathbf{X}_j
 - choisir l'attribut \mathbf{X}_j qui maximise $I_S(\mathbf{X}_j, \mathbf{Y})$
 - créer un **nœud** dans l'arbre de décision avec \mathbf{X}_j
 - partitionner** \mathcal{E} en sous-ensembles avec les valeurs de \mathbf{X}_j
 - mettre les sous-ensembles obtenus dans \mathcal{P}

Marsala & Guigue – 2022

LU3IN026 – cours 7 – 4

1 – Apprentissage par arbres de décision (suite) –

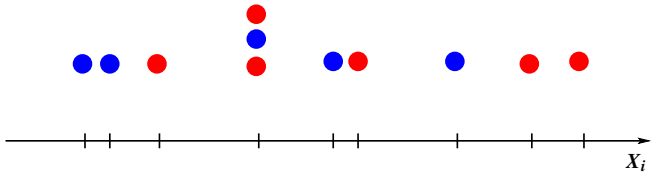
Arbres de décision et données numériques

- Comment faire si l'attribut \mathbf{X}_j est un attribut numérique ?
- Il faut **discrétiser** \mathbf{X}_j : le transformer en attribut **catégoriel**
 - version la plus simple : discrétiser en 2 valeurs catégorielles
- Idée : utilisation d'une valeur de coupure v_j pour \mathbf{X}_j
 - construction de 2 intervalles : $] -\infty, v_j[$ et $[v_j, +\infty[$
 - on note : $\{\mathbf{X}_j, v_j\}$ cette décomposition
- Déterminer la valeur v_j qui minimise $H(C|\{\mathbf{X}_j, v_j\})$
 - essayer toutes les valeurs possibles
- Modification de l'algorithme de construction d'arbre :
 - phase de **discrétisation** de \mathbf{X}_j
 - traiter \mathbf{X}_j comme un attribut catégoriel : $\{\mathbf{X}_j, v_j\}$
- Deux possibilités
 - la discrétisation est faite avant la construction de l'arbre
 - la **discrétisation est faite localement**

Marsala & Guigue – 2022

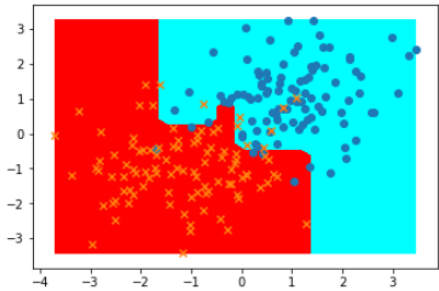
LU3IN026 – cours 7 – 6

Arbres de décision et données numériques (2)

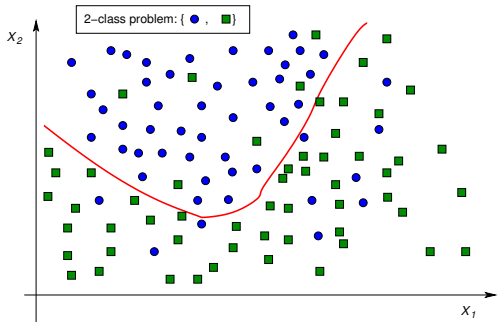


Arbres de décision et données numériques (4)

► Exemple un peu moins simple



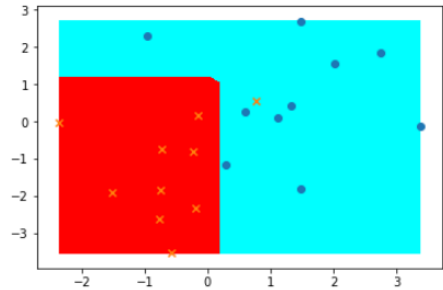
Apprentissage supervisé : séparateur des classes



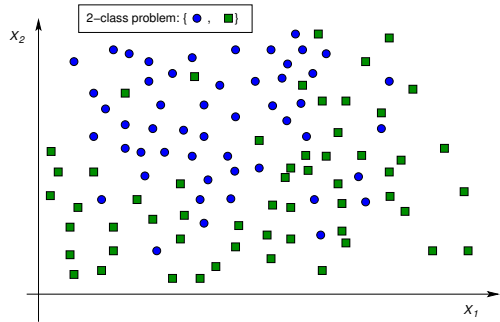
► La **frontière** de décision entre les classes doit être trouvée

Arbres de décision et données numériques (3)

► Exemple simple

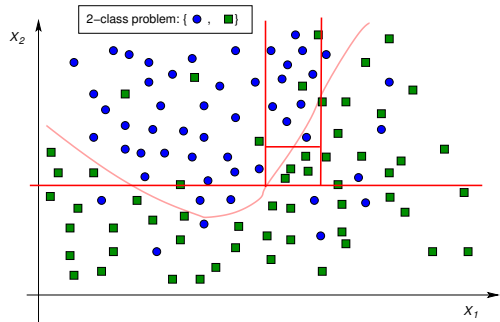


Représentation d'une base d'apprentissage



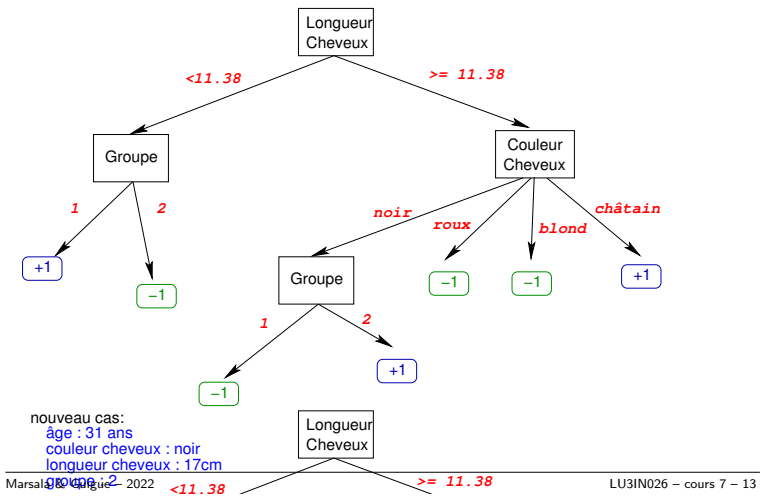
► Étant donné une base d'apprentissage
• descriptions + classes

Frontières fournies par un arbre de décision



► Un arbre de décision définit un découpage par des frontières perpendiculaires aux axes
• chaque frontière est définie par un test d'un nœud de l'arbre

Classification avec un arbre de décision



Plan du cours

Apprentissage par arbres de décision (suite)

Méthodes d'ensembles

Conclusion sur les arbres de décision

- Avantages
 - modèle d'apprentissage **interprétable**
 - mécanismes simples de construction
 - hiérarchie des attributs simple à comprendre
 - utilisation en classification
- Inconvénients
 - frontière construite par coupures perpendiculaires aux axes
 - pas de prise en compte de combinaisons d'attributs possibles
 - sous-apprentissage possible si le critère d'arrêt est trop lâche
 - sur-apprentissage si le critère d'arrêt est trop fort
 - lors de la construction
 - optimisation **locale** pour le choix d'un attribut

Biais et Variance (1)

- Apprentissage : trouver f , fonction de prédiction, telle que :

$$y = f(\mathbf{x}) + \epsilon$$

avec $\epsilon \geq 0$ le plus petit possible

- idéalement : $\epsilon = 0$ (mais on n'y arrive jamais...)
- la "forme" de f est importante : elle utilise les variables de \mathbf{x}
 - linéaire, quadratique,...
 - arbre de décision
 - ...
- Modèle **parcimonieux** : nombre réduit de variables utilisées,...
 - idée : modèle parcimonieux \implies faible variance
- **Biais** : complexité du modèle
- **Variance** : capacité du modèle à changer si la base d'apprentissage change

Biais et Variance (2)

- Objectif : faible biais & variance faible
 - très difficile d'atteindre les 2... il faut choisir !
- Nouvelle approche : réduire la variance
 - combiner plusieurs classifieurs
 - agréger leur résultats pour améliorer les performances
- Différentes façons de faire
 - on regarde avec les arbres (par exemple)
 - multiplier les arbres pour les combiner ensuite