Apprentissage et classification monotone

Laura Nguyen

LFI

19 juillet 2018

Classification monotone

- Garantie d'une monotonie de la classe par rapport aux valeurs d'attributs
- Concepts sémantiques : préférence, priorité, importance...
- Applications :
  - Prédiction du risque de faillite
  - Evaluation du prix de biens immobiliers
  - Cote de crédit
- Gain en interprétabilité (y compris auprès des experts) et en taux de bonne classification quand le domaine s'y prête   Pazzani et al. (2001)

Exemple de problème de classification monotone

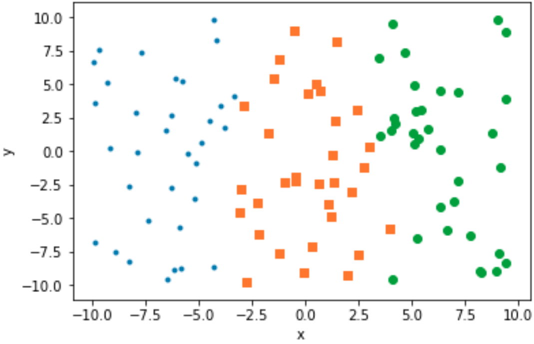
Potharst & Feelders (2002)

client	revenu	éducation	casier judiciaire	prêt
cl1	faible	faible	juste	non
cl2	faible	faible	excellent	faible
cl3	moyen	intermédiaire	excellent	intermédiaire
cl4	élevé	faible	excellent	élevé
cl5	élevé	intermédiaire	excellent	élevé

Formalisation

- Entrées :
  - n exemples : $\Omega = \{\omega_1, ..., \omega_n\}$
  - m attributs ordonnés : $A = (a_1, ..., a_m)$.
    Pour $j = 1, ..., m$:
    - $a_j : \Omega \rightarrow X_j$ avec X_j totalement ordonné
    - Espace de description $X = X_1 \times ... \times X_m$
  - k classes : $C = c_1, ..., c_k$
- Sortie :
  - Fonction de classification monotone $\lambda : X \rightarrow C$

Dataset artificiel à un attribut monotone



Classification par arbres de décision

- Pas d'hypothèse sur les données
- Exploiter au mieux l'éventuelle gradualité
- Classification par arbres de décision
  - Choisir a_j respectant le plus la contrainte de monotonie locale :

    $$\forall \omega_i, \omega_h \in \Omega, a_j(\omega_i) \leq a_j(\omega_h) \Rightarrow \lambda(\omega_i) \leq \lambda(\omega_h)$$

  - Pas de garantie d'avoir un classifieur globalement monotone

Mesures de discrimination à rang

- Problème : incapacité des mesures de discrimination classiques à détecter la monotonie
- Chercher des mesures de discrimination sensibles à la monotonie et robustes au bruit non-monotone
  - Généralisation à rang de mesures classiques (Shannon, Gini)
  - Modèle de construction hiérarchique

Ensembles dominants

- Classes d'équivalence générés par, :
  - un attribut a_j : $[\omega_i]_{a_j} = \{\omega_h \in \Omega : a_j(\omega_i) = a_j(\omega_h)\}$
  - la classe λ : $[\omega_i]_\lambda = \{\omega_h \in \Omega : \lambda(\omega_i) = \lambda(\omega_h)\}$
- Ensembles dominants générés par, :
  - un attribut a_j : $[\omega_i]_{a_j}^{\leq} = \{\omega_h \in \Omega : a_j(\omega_i) \leq a_j(\omega_h)\}$
  - la classe λ : $[\omega_i]_\lambda^{\leq} = \{\omega_h \in \Omega : \lambda(\omega_i) \leq \lambda(\omega_h)\}$

Généralisation à rang de l'entropie de Shannon

- Entropie conditionnelle de Shannon :
  $H_s(\lambda|a_j) = \sum_{i=1}^{|\Omega|} \frac{1}{|\Omega|}(\log(\frac{|[\omega_i]_\lambda \cap [\omega_i]_{a_j}|}{|[\omega_i]_{a_j}|}))$
- Entropie de Shannon à rang :
  $H_s^*(\lambda|a_j) = \sum_{i=1}^{|\Omega|} \frac{1}{|\Omega|}(\log(\frac{|[\omega_i]_\lambda^\leq \cap [\omega_i]_{a_j}^\leq|}{|[\omega_i]_{a_j}^\leq|}))$

Modèle de construction hiérarchique de mesures de discrimination à rang **marsala-rank**

- Isoler les propriétés d'une telle mesure
- Créer de nouvelles mesures
- Structure fonctionnelle commune avec 3 fonctions
  Pour a_j fixé :
  - f^* : mesure de monotonie locale de l'objet
  - g^* : mesure de non-monotonie locale de l'objet
  - h^* : agrégation de mesures de non-monotonie locale

  Ecriture générique :

  $$H^*(\lambda|a_j) = h^*(g^*(f^*(\omega_1)), ..., g^*(f^*(\omega_n)))$$

Couche f^*

Pour $a_j \in A$ fixé,

- $dsr(\omega_i) = \dfrac{|[\omega_i]^{\leq}_{\lambda} \cap [\omega_i]^{\leq}_{a_j}|}{|[\omega_i]^{\leq}_{a_j}|}$
- $mindsr(\omega_i) = \dfrac{min_{\omega_h \in [\omega_i]_{a_j}} |[\omega_h]^{\leq}_{\lambda} \cap [\omega_h]^{\leq}_{a_j}|}{|[\omega_i]^{\leq}_{a_j}|}$
- $maxdsr(\omega_i) = \dfrac{max_{\omega_h \in [\omega_i]_{a_j}} |[\omega_h]^{\leq}_{\lambda} \cap [\omega_h]^{\leq}_{a_j}|}{|[\omega_i]^{\leq}_{a_j}|}$
- $avgdsr(\omega_i) = \dfrac{\displaystyle\sum_{\omega_h \in [\omega_i]_{a_j}} \dfrac{|[\omega_h]^{\leq}_{\lambda} \cap [\omega_h]^{\leq}_{a_j}|}{|[\omega_i]_{a_j}|}}{|[\omega_i]^{\leq}_{a_j}|}$

Réécriture des mesures de discrimination à rang

- $H_S^*(\lambda|a_j) = \sum_{i=1}^{n} \frac{1}{n}(-\log(dsr(\omega_i)))$
  - $f_S^* = dsr(\omega_i)$
  - $g_S^* = -\log(f_S^*(\omega_i))$
  - $h_S^* = \sum_{i=1}^{n} \frac{1}{n} g_S^*(f_S^*(\omega_i))$
- $H_P^*(\lambda|a_j) = \sum_{i=1}^{n} \frac{1}{n}(-\frac{\log(mindsr(\omega_i))}{mindsr(\omega_i)})$
  - $f_P^* = mindsr(\omega_i)$
  - $g_P^* = -\frac{\log f_P^*(\omega_i)}{f_P^*(\omega_i)}$
  - $h_P^* = \sum_{i=1}^{n} \frac{1}{n} g_P^*(f_P^*(\omega_i))$

Construction d'arbres de décision monotones

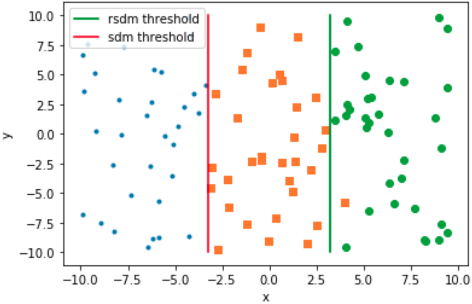
- Classifieur RDMT(H) paramétré par :
  - Une mesure de discrimination H
  - 3 critères de pré-élagage : partitionnement, arrêt, étiquetage
- Critère de partitionnement (splitting rule)
  - Chaque attribut a_j est binarisé : $\forall x \in X_j$,

$$a_j^x(\omega_i) = \begin{cases} 0 & \text{if } a_j(\omega_i) \leq x \\ 1 & \text{otherwise} \end{cases}$$

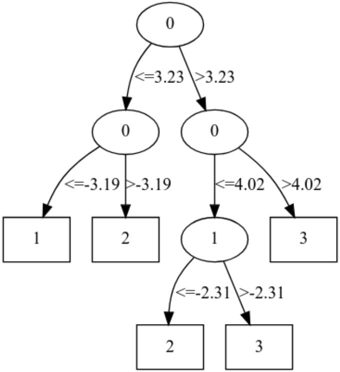
  - Trouver $a_*^{x_*}$ minimisant $H(\lambda | a_j^x)$
    - ★ a_* est l'attribut utilisé pour le partionnement
    - ★ x_* est la valeur seuil :

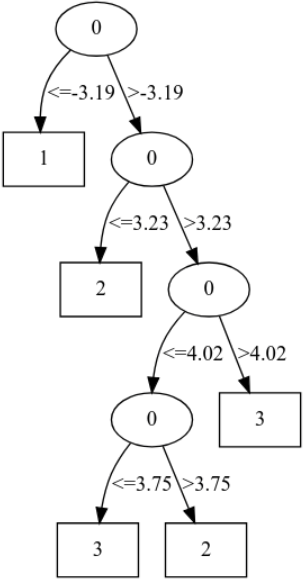
$$x_* = arg\ min\{H(\lambda | a^x) | j = 1, ..., m, x \in X_j\}$$

Critère de partitionnement



Arbres de décision générés





RDMT(H) sur des datasets artificiels

RDMT(H) sur de vrais datasets

Pazzani, M. J., Mani, S. & Shankle, W. R. (2001). Acceptance of rules generated by machine learning among medical experts. In *Methods of information in medicine* 40.05, p. 380-385.

Potharst, R. & Feelders, A. J. (2002). Classification trees for problems with monotonicity constraints. In *ACM SIGKDD Explorations Newsletter* 4.1, p. 1-10.