

APPRENTISSAGE: APPRENTISSAGE DE RÉSEAUX BAYÉSIENS

M. Serrurier
IRIT, Toulouse, France

December 8, 2015

- ▶ Prédiction en termes de probabilité : Prédit plusieurs hypothèses en les pondérant par leurs probabilités
- ▶ Etant donné un objet O , la méthode consiste à calculer la probabilité d'appartenance de O à chaque classe, puis choisir celle qui maximise cette valeur
- ▶ Standard : Même s'il s'avère que les méthodes bayésiennes se révèlent impossible à calculer, elles peuvent être considérées comme étalon pour mesurer la correction d'autres méthodes

- ▶ Approche probabiliste basé sur les probabilités conditionnelles
- ▶ Le problème de classification peut être formalisé en utilisant les probabilités a-posteriori :
- ▶ $P(C|X) = \text{prob. que } X = \langle x_1, \dots, x_k \rangle \text{ soit de la classe } C.$

Principe

Affecter à X la classe C tel que $P(C|X)$ est maximal

Théoreme de Bayes

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

- ▶ $P(X)$ est la même pour toutes les classes
- ▶ $P(C)$ = fréquence relative des éléments de C
- ▶ C telle que $P(C|X)$ est maximum = C telle que $P(X|C) \cdot P(C)$ est maximum
- ▶ Problème : calculer $P(X|C)$ est infaisable !

Exercice

Si X est décrit par 32 variables binaires, combien de probabilités doit on calculer.

ESTIMER LES PROBABILITÉS A-POSTERIORI

// QUELQUES OBSERVATIONS

- ▶ $P(C|X) = P(X|C) \cdot P(C) / P(X)$
 - ▶ $P(C|X)$ croît quand $P(C)$ croît : plus C est probable, plus il y a de chances qu'elle soit la classe
 - ▶ $P(C|X)$ croît quand $P(X|C)$ croît : si X arrive souvent quand C est la classe, alors il y a des chances que C soit la classe
 - ▶ $P(C|X)$ décroît quand $P(X)$ croît : si X est courant, il nous apprend peu sur C

Hypothèse Maximale A Posteriori

$$h_{\text{map}} = \operatorname{argmax}_{c_k \in C} \frac{P(X|c_k) \cdot P(C)}{P(X)}$$

- ▶ si $P(X)$ est la même pour toutes les classes alors

$$h_{\text{map}} = \operatorname{argmax}_{c_k \in C} P(X|c_k) \cdot P(C)$$

Maximum de vraisemblance

$$h_{\text{ML}} = \operatorname{argmax}_{c_k \in C} P(X|c_k)$$

- ▶ Appliquer le théorème de Bayes pour définir un algorithme de classification simple et efficace (en pratique)
- ▶ Caractéristiques :
 - ▶ Classification supervisée (mais peut être étendu au cas non supervisé)
 - ▶ Entrée discrètes et continues
 - ▶ Classes discrètes
- ▶ But : contourner le problème du calcul de $P(X|C)$.

Hypothèse

indépendance des attributs

$$P(x_1, \dots, x_q | C) = P(x_1 | C) \cdot \dots \cdot P(x_q | C)$$

- ▶ Si attribut A_i est qualitatif : $P(x_i | C) = \frac{n_{ic}}{n_c}$ est estimée par la fréquence relative des éléments ayant la valeur x_i pour A_i et qui sont dans C (n_{ic}) par rapport aux éléments dans C (n_c).
- ▶ Si attribut A_i est continu : $P(x_i | C)$ peut être estimé comme une loi normale (par exemple)
- ▶ Facile à calculer à partir des exemples dans les deux cas

- ▶ Attributs binaires : $2 * q$ valeurs à estimer
- ▶ Naïf : hypothèse d'indépendance (jamais vérifiée, simplification)
- ▶ Procédure sub-optimale, mais intérêt pratique
- ▶ Nombreuses applications comme la classification de texte.

CLASSIFIEUR BAYÉSIEEN NAÏF

// EXERCICE 1

saire	impôts	étudiant	contrôle
< 30	< 20 %	oui	négatif
30 - 50	< 20 %	non	positif
30 - 50	< 20 %	oui	positif
30 - 50	> 20 %	non	négatif
> 50	< 20 %	non	positif

Question

Faut-il faire un contrôle fiscale à 35, 6%, oui ?

CLASSIFIEUR BAYÉSIEEN NAÏF

// EXERCICE 2

Ciel	Température	Humidité	vent	Jouer
Soleil	Chaud	Forte	faible	Non
Soleil	Chaud	Forte	Fort	Non
Couvert	Chaud	Forte	faible	Oui
Pluie	Doux	Forte	faible	Oui
Pluie	Frais	Normale	faible	Oui
Pluie	Frais	Normale	Fort	Non
Couvert	Frais	Normale	Fort	Oui
Soleil	Doux	Forte	faible	Non
Soleil	Frais	Normale	faible	Oui
Pluie	Doux	Normale	faible	Oui
Soleil	Doux	Normale	Fort	Oui
Couvert	Doux	Forte	Fort	Oui
Couvert	Chaud	Normale	faible	Oui
Pluie	Doux	Forte	Fort	Non

Question

Quelle classe attribuer à : (Soleil,Frais,Forte,Fort) ? Calculer la probabilité d'être dans la classe choisie

- ▶ Pas de construction de modèle
- ▶ Pas d'explication
- ▶ **Prédiction** : comparable aux autres algorithmes
- ▶ **Vitesse** : linéaire en apprentissage, constant en classification
- ▶ **Robustesse (bruit)** : très bonne
- ▶ **Changement d'échelle** : les exemples ne sont pas en mémoire centrale (occupation mémoire = nb attributs x nb classes)

- ▶ Rend le calcul possible
- ▶ Problème : en pratique, les attributs (variables) sont souvent corrélés

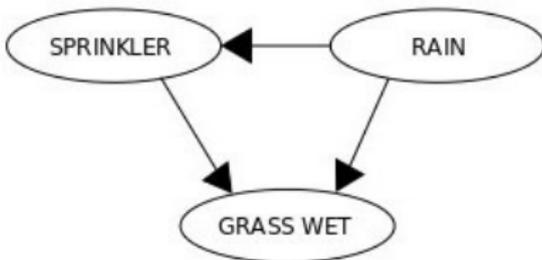
Solution

Réseaux Bayésiens, utiliser le raisonnement Bayésien en tenant compte des relations causales qui existent entre attributs

RÉSEAUX BAYÉSIENS

// EXEMPLES

RAIN	SPRINKLER	
	T	F
F	0.4	0.6
T	0.01	0.99



RAIN	T	F
	0.2	0.8

SPRINKLER	RAIN	GRASS WET	
		T	F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

- ▶ Un tel réseau autorise un sous ensemble d'attributs indépendants (ex : Historique familial et Fumeur)
- ▶ Chaque nœud dépend de ces antécédents.
- ▶ Le modèle graphique représente les relations causales
- ▶ La table des probabilités conditionnelles d'une variable tient compte de toutes les combinaisons possibles de ses antécédents
- ▶ Soit $X = \langle x_1, \dots, x_n \rangle$ un tuple. $\text{Prob}(X)$ est donné par

$$\prod_i P(x_i | \text{Parents}(X_i))$$

- ▶ Cas le plus simple
- ▶ Si la table est complète :
 - ▶ Il suffit de calculer les fréquences des tables conditionnelles
- ▶ Si la table est incomplète :
 - ▶ On estime les probabilité conditionnelles avec un algorithme d'optimisation (EM par exemple)
 - ▶ Peut être long si le réseau est compliqué

- ▶ Basée sur une mesure de pertinence entre un réseau et les données
- ▶ On suppose qu'étant donnée une structure on peut calculer les tables
- ▶ Approche par optimisation :
 - ▶ On part d'un réseau au hasard
 - ▶ Tant que ça ne converge pas
 - ▶ On choisit une perturbation du réseau au hasard (ajouté ou retirer un arc)
 - ▶ Si le nouveau réseau améliore le score, on le garde
- ▶ Peut être très coûteux en temps de calcul

- ▶ Avantages :
 - ▶ Peu prédire n'importe quel attribut ou combinaison d'attributs
 - ▶ Représente la dépendance entre les variables
- ▶ Inconvénients :
 - ▶ Difficiles à apprendre
 - ▶ Marche mal avec les données numériques