Université de Montréal
Département d'informatique et de recherche opérationnelle

IFT3395/6390
Fondements de l’apprentissage machine

Classifieur de Bayes.
Classifieur de Bayes Naïf

Professeur: Pascal Vincent

LISA Laboratoire d’Informatique des Systèmes d’Apprentissage

Approche probabiliste de l'apprentissage

- On suppose que les données sont générées par un processus inconnu.
- X,Y est vu comme une paire de variables aléatoires, distribuées selon une loi de probabilité inconnue P(X,Y).
- X (une variable vectorielle) est elle-même vue comme un ensemble de variables aléatoires scalaires.

$$P(X,Y) = P(X_{[1]}, \ldots, X_{[d]}, Y)$$

Classifieur de Bayes

ou comment construire un classifieur à partir d'estimations de densité

- On sépare l'ensemble d'entraînement en m sous-ensembles contenant chacun tous les points d'une même classe.

- On entraîne un estimateur de densité sur chacun: $c \in \{1, .., m\}$

  $$\hat{p}_c(x) \simeq p_{X|Y}(x|c)$$

- On détermine les probabilités à priori de chaque classe $\hat{P}_c = \frac{n_c}{n} \simeq P_Y(c) = P(Y = c)$
  (par ex. en comptant leurs proportions relatives dans l'ensemble d'apprentissage)

- On applique la règle de Bayes pour obtenir la probabilité à postériori des classes au point x.

- On choisit la plus probable.

$$\overbrace{P_{Y|X}(c|x)}^{\textit{posterior}} = \frac{\overbrace{p_{X|Y}(x|c)}^{\substack{\textit{class-conditional density} \\ \textit{(likelihood)}}} \overbrace{P_Y(c)}^{\textit{prior}}}{p_X(x)}$$

$$= \frac{p_{X|Y}(x|c) P_Y(c)}{\sum_{c'=1}^{m} p_{X|Y}(x|c') P_Y(c')}$$

$$\simeq \frac{\hat{p}_c(x) \hat{P}_c}{\sum_{c'=1}^{m} \hat{p}_{c'}(x) \hat{P}_{c'}}$$

Classifieur de Bayes Naïf

- Dans le classifieur de Bayes Naïf, on suppose, pour chaque classe $c \in \{1, .., m\}$ que, étant donné c, les composantes de X sont indépendantes:

$$\begin{aligned} P(X|Y = c) &= P(X_{[1]}|Y = c)P(X_{[2]}|Y = c)...P(X_{[d]}|Y = c) \\ \hat{p}_c(x) &= \hat{p}_{c,1}(x_{[1]})\hat{p}_{c,2}(x_{[2]})...\hat{p}_{c,d}(x_{[d]}) \end{aligned}$$

- Il suffit donc de modéliser des densités univariées, les $\hat{p}_{c,j}(x_{[j]})$ ce qui est une tâche facile (univariée == dimension 1: pas de fléau de la dimensionalité; les méthodes de type histogramme ou Parzen fonctionnent plutôt bien).

- On construit ensuite un classifieur de Bayes à partir des estimateurs $\hat{p}_c(x)$ ainsi obtenus.

Attention

- “Classifieur de Bayes” $\neq$ “Classifieur de Bayes naif”
- “Classifieur de Bayes naif” $\subset$ “Classifieur de Bayes”
- Classifieur de Bayes naif:
  modèle très restrictif, de faible capacité
  (ne peut pas tenir compte des interactions entre les composantes)

Autre façon de construire un classifieur: à partir d'un bon estimateur de la probabilité

- On estime la probabilité jointe P(X,Y)
- On peut calculer la probabilité conditionnelle de la classe c:

$$
\begin{aligned}
P(Y=c|X=x) &= \frac{P(X=x,Y=c)}{P(X=x)} \\
&= \frac{P(X=x,Y=c)}{\sum_{c'=1}^{m} P(X=x,Y=c')}
\end{aligned}
$$

- Notez que les proba de classe (conditionelles à x) sont proportionnelles aux probas jointes. Le dénominateur est une simple normalisation pour qu'elles somment à 1.