APV - TP5 : Naive Bayes Classifier   Hugo Huart

Exercice 1

On a $h_\theta(x) = \arg\min_{c_j \in Y} \sum_{i=1}^{m} L_j(c_i)\, p(c_i|x)$

et on sait que $L_j(c_j) = 0$ et $L_j(c_i) = 1$ si $i \neq j$.

Avec la formule de Bayes : $P(A|B) = \dfrac{P(B|A) \cdot P(A)}{P(B)}$

on peut effectuer le développement suivant :

$$h_\theta(x) = \arg\min_{c_j \in Y} \sum_{i=1}^{m} L_j(c_i)\, p(c_i|x)$$

$$= \arg\min_{c_j \in Y} \left(\sum_{i=1}^{m} p(c_i|x) \right) - p(c_j|x)$$

$$= \arg\min_{c_j \in Y} \left(\sum_{i=1}^{m} \frac{p(x|c_i) \cdot p(c_i)}{p(x)} \right) - p(c_j|x)$$

$$= \arg\min_{c_j \in Y} \frac{1}{p(x)} \left(\sum_{i=1}^{m} p(x|c_i) p(c_i) \right) - p(c_j|x)$$

$$= \arg\min_{c_j \in Y} \frac{1}{p(x)} \left(\sum_{i=1}^{m} p(x \wedge c_i) \right) - p(c_j|x)$$

$$= \arg\min_{c_j \in Y} \frac{1}{p(x)} p(x) - p(c_j|x)$$

$$= \arg\min_{c_j \in Y} - p(c_j|x)$$

$$= \arg\max_{c_j \in Y} p(c_j|x)$$

$$= \arg\max_{c_j \in Y} \frac{p(x|c_j) p(c_j)}{p(x)}$$

$$= \underline{\underline{\arg\max_{c_j \in Y} p(x|c_j) p(c_j)}}$$

Exercice 2

On a le modèle suivant :

$$h_\theta(x) = \arg\max_{c_j \in Y} \prod_{i=1}^{k} p(x_i | c_j)\, p(c_j)$$

À l'aide des données d'entraînement on peut calculer les probabilités empiriques suivantes,

pour C_V et C_S les classes respectives des iris Versicolore et Setosa.

$$p(C_V) = 6/12 = \frac{1}{2}$$

$$p(C_S) = 6/12 = \frac{1}{2}$$

On calcule maintenant les probabilités empirique pour chaque attribut de notre observation et pour chaque classe.

Versicolores :

$p(\text{SépaleVerte} | C_v) = \frac{2}{6} = \frac{1}{3}$

$p(\text{SépaleArrondie} | C_v) = \frac{1}{6}$

$p(\text{Pétale Jaune} | C_v) = \frac{2}{6} = \frac{1}{3}$

$p(\text{Pétale Petite} | C_v) = \frac{1}{6}$

Setosa :

$p(\text{Sépale Verte} | C_s) = \frac{2}{6} = \frac{1}{3}$

$p(\text{Sépale Arrondie} | C_s) = \frac{4}{6} = \frac{2}{3}$

$p(\text{Pétale Jaune} | C_s) = \frac{4}{6} = \frac{2}{3}$

$p(\text{Pétale Petite} \mid C_S) = \frac{4}{6} = \frac{2}{3}$

On a donc :

$$\prod_{i=1}^{k} p(x_i \mid C_V)\, p(C_V) = \frac{1}{3} \cdot \frac{1}{6} \cdot \frac{1}{3} \cdot \frac{1}{6} \cdot \frac{1}{2} \simeq 0.00154$$

$$\prod_{i=1}^{k} p(x_i \mid C_S)\, p(C_S) = \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{2} \simeq \underline{\underline{0.0494}}$$

On a $\operatorname{argmax}_{C_j \in Y} \prod_{i=1}^{k} p(x_i \mid C_j)\, p(C_j) = C_S$,

on estime alors que notre fleur est une

<u>iris setosa.</u>