Devoir #1 - IFT3395/6390

Julien Allard et André Langlais

12 Octobre 2017

1 Petit exercice de probabilités

Soit les variables aléatoires suivantes:
X: Femme atteinte du cancer
Y: Test de détection est positif

Formule de Bayes:

$$P(X|Y) = \frac{P(X) * P(Y|X)}{P(Y)} \tag{1}$$

Les données de la question nous permettent de trouver les probabilités suivantes:

$$\begin{gathered} P(X) = 0.01 \\ P(Y|X) = 0.8 \end{gathered} \tag{2}$$

Il nous manque à trouver P(Y) qui peut être calculer par la formule suivante:

$$P(Y) = P(Y|X) + P(Y|X') \tag{3}$$

où X' répresente l'évènement complémentaire de X

$$P(Y) = 0.01 * 0.80 + 0.99 * 0.096 = 0.10304 \tag{4}$$

Nous avons maintenant toutes les informations requises pour appliquer le théorème de Bayes

$$P(X|Y) = \frac{0.01 * 0.80}{0.10304} \approx 0.0774 \tag{5}$$

La bonne réponse est donc 6. soit moins que 10%. Il peut paraître surprenant que le pourcentage soit si bas, mais on peut vérifier visuellement pourquoi on obtient un tel pourcentage

2 Estimation de densité : paramétrique Gaussienne, v.s. fenêtres de Parzen

2.1 Gaussienne isotropique

(a)

Les paramètres sont la moyenne μ de dimension d et la variance σ de dimension 1

(b)

$$\begin{aligned} n &= |D| \\ \mu_{MaxVraiss} &= \sum_{i}^{n} \frac{x_i}{nd} \\ \sigma^2_{MaxVraiss} &= \sum_{i}^{n} \frac{(x_i - \mu)^T (x_i - \mu)}{nd} \end{aligned} \tag{6}$$

(c)

μ se calcule en $O(nd)$ et σ^2 se calcule en $O(nd)$ aussi

(d)

$$p_{gauss-isotrop}(x) = \frac{1}{(2\pi)^{\frac{d}{2}} \sigma^d} \exp^{\frac{-1}{2} \frac{\|x-\mu\|^2}{\sigma^2}} \tag{7}$$

(e)

Le temps de calcul pour la prédiction d'un point de test x se fait en $O(1)$

2.2 Fenêtres de Parzen avec noyau Gaussien isotropique

(a)

Puisque σ est fixé, la phase d'entraînement consiste à garder l'ensemble d'entraînement en mémoire.

(b)

$$p_{mathrmParzen}(x) = \frac{1}{n} \sum_{i}^{n} \frac{1}{(2\pi)^{\frac{d}{2}} \sigma^d} \exp^{\frac{-1}{2} \frac{\|x-x_i\|^2}{\sigma^2}} \tag{8}$$

(c)

Le temps de calcul pour la prédiction d'un point de test x se fait en $O(n)$

2.3 Capacité

(a)

Parzen a plus de capacité puisqu'on peut définir des régions de décisions plus précise. Il y a plus de changement rapide lorsque l'on se déplace dans la région totale.

(b)

Parzen est le modèle qui peut plus facilement être dans un cas de surapprentissage. Cela se produit dans le cas où les données sont plus éparpillées et lorsque le *sigma* est très petit.

(c)

Pour la Gaussienne, on veut optimiser les paramètres selon les données. Elles sont optimisables avec la méthode de vraisemblance. Pour Parzen, on ne peut pas optimiser le *sigma* selon la méthode de vraisemblance. En effet, la variance de la Gaussienne au point x ne se calcule pas puisqu'il n'y a qu'une seule donnée pour cette Gaussienne.

2.4 Gaussienne diagonale

(a)

$$p_{gausDiag}(x) = \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{|\Sigma|}} \exp^{\frac{-1}{2}\|x-\mu\|^T\Sigma^{-1}(x-\mu)} \tag{9}$$

(b)

Les composantes sont chacune tirées d'un distribution gaussienne donc elles sont aléatoires. De plus, le fait de tirer une des composantes n'impacte pas les prochaines donc elles sont indépendantes.

(c)

Voici l'équation qui permettrait de minimiser le risque empirique selon le maximum de vraisemblance

$$\prod_i^n p(x_i) = -log(\prod_i^n \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{|\Sigma|}} \exp^{\frac{-1}{2}\|x_i-\mu\|^T\Sigma^{-1}(x_i-\mu)}) \tag{10}$$

(d)

Résolution de l'équation

$$
\begin{aligned}
\prod_i^n p(x_i) &= -\sum log(\frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{|\Sigma|}} \exp^{\frac{-1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)}) \\
&= -\sum log(\frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{|\Sigma|}}) - \frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu) \\
&= -\sum -log(2\pi) - log(\sqrt{|\Sigma|}) - \frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu) \\
&= \frac{nd}{2}log(2\pi) + nlog(\sqrt{|\Sigma|}) + \sum \frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)
\end{aligned}
\tag{11}
$$

On dérive par rapport à μ pour trouver le premier paramètre

$$
\frac{\partial}{\partial\mu}\prod_i^n p(x_i) = 0 + 0 + \sum 2(\mu - x_i)\Sigma^{-1} \tag{12}
$$

On la met égal à zéro

$$
\begin{aligned}
\frac{\partial}{\partial\mu}\prod_i^n p(x_i) &= 0 \\
\sum 2(\mu - x_i)\Sigma^{-1} &= 0 \\
\mu &= \frac{\sum_i^n x_i}{n}
\end{aligned}
\tag{13}
$$

2.5 Problème de classification

(a)

En se basant sur un classifieur de Bayes et en utilisant comme modèle une Gaussien isotropique, on doit d'abord entraîner les données en estimant μ et σ^2 à l'aide de la méthode de vraisemblance. Ensuite, on peut déterminer les probabilités à priori pour les différentes classes.

(b)

$$
\begin{aligned}
& c \in \{1, ..., n\} \\
B_{Gaussisotrop}(x) &= \frac{1}{(2\pi)^{\frac{d}{2}}\sigma^d} \exp^{\frac{-1}{2}\frac{\|x-\mu\|^2}{\sigma^2}} + \log(P_c = \frac{n_c}{n} \approx P(Y = c))
\end{aligned}
\tag{14}
$$