
EXAMEN DE CLASSIFICATION - 3TSI

Vendredi 25 octobre 2019

Polycopié de cours et transparents autorisés

Exercice 1 : Classification Bayésienne (6 points)

On considère un problème de classification à deux classes ω_1 et ω_2 de densités uniformes sur les intervalles $[0, a]$ et $[0, 1]$ avec $a < 1$, c'est-à-dire

$$f(x|\omega_1) = \frac{1}{a}\mathcal{I}_{[0,a]}(x) \quad \text{et} \quad f(x|\omega_2) = \mathcal{I}_{[0,1]}(x) \quad (1)$$

où \mathcal{I}_A est la fonction indicatrice sur l'ensemble A telle que $\mathcal{I}_A(x) = 1$ si $x \in A$ et $\mathcal{I}_A(x) = 0$ si $x \notin A$. On notera $P(\omega_1) = p \neq \frac{1}{2}$.

1. (2 pts) Déterminer les probabilités $P(\omega_1|x)$ et $P(\omega_2|x)$ et représenter les graphiquement (en fonction de x) pour $p = 1/4$ et $p = 3/4$, en distinguant avec soin les deux cas $x \in [0, a]$ et $x \in]a, 1]$.
2. (2 pts) Déterminer la règle de classification associée à ce problème avec la fonction de coût 0 – 1 dans les deux cas $p = \frac{1}{4}$ et $p = \frac{3}{4}$.
3. (2 pts) Déterminer la probabilité d'erreur associée à ce classifieur en fonction de a et p , en distinguant les deux cas $p < \frac{1}{2}$ et $p \geq \frac{1}{2}$.

Exercice 2 : Détection d'anomalies (4 points)

La méthode one-class SVM détermine un hyperplan séparateur par résolution du problème d'optimisation suivant

$$\begin{aligned} \underset{\mathbf{w}, \rho, \xi_i}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{w}\|^2 - \rho + \frac{1}{\nu N} \sum_{i=1}^N \xi_i \\ \text{with} \quad & \langle \mathbf{w}, \mathbf{x}_i \rangle \geq \rho - \xi_i, \forall i = 1, \dots, N. \end{aligned}$$

1. (1pt) Quelle est la règle de décision permettant de décider si un vecteur test \mathbf{x} est une anomalie.
2. (1pt) Quel est le rôle des variables latentes ξ_i ?
3. (1pt) Comment règle-t-on le paramètre ν et à quoi correspond-il ?
4. (1pt) Dans le cas d'un noyau gaussien défini par $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2)$, comment règle-t-on le paramètre γ ?

Questions sur l'article (10 points)

1. (1 pt) Expliquer la phrase "This is a supervised learning problem" utilisé dans l'introduction de la section 2.
2. (1 pt) Expliquer comment un document d est transformé en un vecteur de paramètres.
3. (2 pts) Quelle est la définition du gain en information utilisé par les auteurs pour déterminer les mots les plus discriminants ? Expliquer comment chaque terme de ce gain en information peut être calculé.
4. (1 pt) Expliquer pourquoi les vecteurs de données sont parcimonieux (sparse).
5. (2 pts) Expliquer avec soin comment on détermine le vecteur \vec{w} de la règle de décision (4) à partir des données d'apprentissage dans le cas de deux classes (i.e., ici de deux catégories).
6. (1 pt) Dans le cas d'utilisation d'un noyau RBF (radial basis function), comment les auteurs proposent-ils de déterminer γ ?
7. (1 pt) Rappeler le principe du classifieur Bayésien évoqué dans la section 4.1.
8. (1 pt) Expliquer le principe de l'algorithme de classification C4.5.