IML

Examen 2018/2019

- Durée : 2h
- *Documents autorisés : cours, notes de cours et calculatrice*

1 Relay (5 points)

Soit un problème de classification à deux classes dont les données sont $\{(x_i, y_i) \in \mathbb{R} \times \{\mathcal{C}_1, \mathcal{C}_2\}\}_{i=1}^N$ On suppose que pour la classe $\mathcal{C}_1$ x suit une loi normale $p(x|\mathcal{C}_1) = \mathcal{N}(\mu, \sigma^2)$ et pour $\mathcal{C}_2$ x suit suit une loi de Rayleigh $p(x|\mathcal{C}_2) = \dfrac{x}{\lambda^2} \exp^{-\frac{x^2}{\lambda^2}}$.

On suppose que les deux classes sont équiprobables.

1. On veut déterminer les paramètres de la loi de distribution conditionnelle des classes 1 et 2. Estimer les paramètres $\hat{\lambda}$, $\hat{\mu}$ et $\hat{\sigma}$ au sens du maximum de vraisemblance.

   Pour classer les données avec l'approche de Bayes, on prend des coûts 0-1 suivants.

2. Donner l'expression des risques conditionnels $R(\mathcal{C}_k|x)$. Sous quelle condition sur $P(\mathcal{C}_1|x)$ et $P(\mathcal{C}_2|\mathbf{x})$ on décide d'affecter $\mathbf{x}$ à la classe $\mathcal{C}_1$?
3. Donner l'expression de la fonction de décision en fonction de x, des paramètres $\hat{\mu}, \hat{\sigma}$ et $\hat{\lambda}$.

2 Un Nu nommé SVM (11 points)

Soit un ensemble de données $\mathcal{D} = \mathcal{D}_+ \cup \mathcal{D}_-$ avec $\mathcal{D}_+ = \{(\mathbf{x}_i, y_i = 1)\}_{i=1}^m$ l'ensemble des points avec le label $+1$ et $\mathcal{D}_- = \{(\mathbf{x}_i, y_i = -1)\}_{i=m+1}^n$ celui des points de label -1. Chaque point $\mathbf{x} \in \mathbb{R}^d$. On veut apprendre une variante de SVM (appelé ν-SVM) qui pondère différemment les erreurs sur les classes $\mathcal{D}_+$ et $\mathcal{D}_-$. Le modèle recherché est $f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b$ avec $\mathbf{w} \in \mathbb{R}^d$ et $b \in \mathbb{R}$. On veut résoudre le problème suivant

$$
\begin{aligned}
\min_{\mathbf{w}, b, \rho, \{\xi_i\}} \quad & \frac{1}{2}\|w\|^2 - \nu\rho + \frac{\gamma}{n}\sum_{i=1}^m \xi_i + \frac{1-\gamma}{n}\sum_{i=m+1}^n \xi_i \qquad (1)\\
\text{sous les contraintes} \quad & y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq \rho - \xi_i \quad \forall\, i = 1, \cdots, n \\
& \xi_i \geq 0, \quad \forall i = 1, \cdots, n \\
& \rho \geq 0
\end{aligned}
$$

avec ρ la marge qu'on souhaite maximiser. $\nu > 0$ et $0 \leq \gamma \leq 1$ **sont des hyper-paramètres fixés par l'utilisateur**.

1. Donner le nombre de contraintes du problème
2. Montrer que la fonction (1) à minimiser peut s'écrire sous la forme

$$\frac{1}{2}\|w\|^2 - \nu\rho + \sum_{i=1}^n C_i \xi_i$$

   avec l'expression des coefficients C_i à spécifier en fonction de γ, n.

3. Écrire le Lagrangien correspondant à ce problème d'optimisation

4. Donner les conditions d'optimalité par rapport aux variables primales $\mathbf{w}$, b, ξ_k ($\forall k = 1, \cdots, n$) et ρ. En déduire l'expression de $\mathbf{w}$.
5. Formuler alors le problème dual correspondant.
6. Soit le vecteur $\boldsymbol{\alpha}^*$, la solution du problème dual.
   (a) Comment peut-on reconnaître les points sur la marge à partir de $\boldsymbol{\alpha}^*$? Justifer la réponse.
   (b) Connaissant ces points, proposer une facon de calculer b et ρ.
7. Pour sélectionner le modèle optimal quelqu'un propose la procédure suivante : "on choisit plusieurs valeurs de γ, ν et ρ, on apprend un modèle pour chaque triplet (γ, ν, ρ), on l'évalue sur le jeu d'apprentissage et on retient le modèle donnant la meilleure performance en apprentissage". Est-ce que cette procédure vous semble correcte ? Si non que quelle solution vous proposez ?

3 Regroupement hiérarchique (4 points)

Le tableau suivant contient la matrice de distances entre six points `A, B, C, D, E, F.`

	A	B	C	D	E	F
A	0					
B	0.12	0				
C	0.51	0.25	0			
D	.0.84	0.16	0.14	0		
E	0.28	0.77	0.7	0..45	0	
F	0.34	0.61	0.93	0.20	0.67	0

1. Réaliser le clustering de ces données en utilisant la CHA avec une métrique de type *saut minimal*. On dessinera le dendogramme et on indiquera le résultat du clustering pour $K = 2$
2. Répondre à la même question en considérant cette fois une métrique de type *saut maximal*.
3. Pour déterminer le nombre de clusters, on envisage les solutions suivantes :
   — faire une validation croisée,
   — Calculer le critère $J(K) = \sum_{k=1}^{K} \sum_{i \in \mathcal{C}_k} \|\mathbf{x}_i - \mathbf{c}_k\|^2$ pour différentes valeurs de K et choisir le nombre de clusters correspondant au minimum de $J(K)$

   Qu'en pensez-vous ? Justifiez vos réponses.