
EXAMEN DE CLASSIFICATION - 3TSI

Lundi 27 novembre 2017

Polycopié de cours et transparents autorisés

Exercice 1 : Classification Bayésienne pour des lois de Rayleigh (4 points)

On considère un problème de classification à deux classes ω_1 and ω_2 de densités

$$f(x|\omega_i) = \frac{x}{\sigma_i^2} \exp\left(-\frac{x^2}{2\sigma_i^2}\right) I_{\mathbb{R}^+}(x) \quad i = 1, 2 \quad (1)$$

où $I_{\mathbb{R}^+}(x)$ est la fonction indicatrice sur \mathbb{R}^+ ($I_{\mathbb{R}^+}(x) = 1$ si $x > 0$ et $I_{\mathbb{R}^+}(x) = 0$ sinon) et $\sigma_1^2 > \sigma_2^2$.

1. (1.5 pts) Déterminer la règle de classification associée à ce problème avec la fonction de coût 0-1 et lorsque les deux classes sont équiprobables.
2. (1.5 pts) Déterminer la probabilité d'erreur associée à ce classifieur.
3. (1 pt) Si le paramètre σ_1^2 est inconnu, expliquer comment l'estimer à partir de données d'apprentissage de la classe ω_1 en utilisant la méthode du maximum de vraisemblance.

Exercice 2 : Arbres de décision (4 points)

On cherche à construire un arbre de décision permettant de décider si un individu doit jouer au tennis ou non. Une base d'apprentissage a été construite comme suit

	Ciel	Température	Vent	Jouer
x_1	soleil	chaud	faible	Oui
x_2	soleil	chaud	fort	Oui
x_3	couvert	chaud	faible	Non
x_4	pluie	froid	faible	Non
x_5	pluie	froid	faible	Non
x_6	pluie	froid	fort	Oui

1. (1 pt) Déterminer l'indice de Gini associé à cette base d'apprentissage vis-à-vis des deux classes "Jouer au Tennis" et "Ne pas jouer au Tennis".
2. (2pts) Déterminer la variation de l'indice de Gini lorsqu'on coupe les données à l'aide des variables "Ciel", "Température" et "Vent". En déduire la variable qui sera utilisée au premier niveau de l'arbre de décision.
3. (1pt) Expliquer comment on pourrait procéder si la variable "Température" était une valeur en degrés celcius.

Questions sur l'article (12 points)

Remarque : toutes vos réponses doivent être justifiées avec soin.

1. (1 pt) Expliquer pourquoi la matrice \mathbf{X} représentée sur la Fig. 1 est idéalement diagonale par blocs.
2. (1 pt) Expliquer la phrase “It has been shown that learning a dictionary from the training samples instead of using all of them as a dictionary can further enhance the performance of SRC”.
3. (1 pt) Quel est l'intérêt d'ajouter le dictionnaire partagé \mathbf{D}_0 aux dictionnaires des différentes classes $\mathbf{D}_1, \dots, \mathbf{D}_C$?
4. (1 pt) Les auteurs de l'article précisent (bas de la page 5162) qu'avec le dernier terme $\|\mathbf{X}\|_F^2$ la fonction de coût devient convexe par rapport à \mathbf{X} . Quel est l'intérêt d'avoir une fonction de coût convexe ?
5. (1 pt) Expliquer comment les auteurs justifient le fait que le dictionnaire \mathbf{D}_0 doit être de rang faible (voir section II C).
6. (1 pt) Comment est définie la norme nucléaire d'une matrice (“nuclear norm”) ?
7. (1 pt) Expliquer la présence du terme de régularisation $\|\mathbf{X}^0 - \mathbf{M}^0\|_F^2$ dans (4).
8. (1 pt) Expliquer la règle de classification donnée dans (7).
9. (1 pt) Expliquer le principe utilisé pour optimiser la fonction $f_Y(\mathbf{D}, \mathbf{X})$ de (3).
10. (1pt) Qu'est ce qu'on entend par “five-fold cross validation”? (voir Section IV A)
11. (1pt) Expliquer le terme “Random projection matrix” utilisé pour la base de données “Extended YaleB” (voir Section IV A).
12. (1pt) Rappeler le principe de l'analyse en composantes principales utilisée pour les données “AR gender”.
13. Questions Bonus
 - (1pt) Expliquer le principe de l'algorithme “Online Dictionary Learning (ODL)”.
 - (1pt) Pour minimiser la fonction de coût

$$\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \lambda\|\mathbf{x}\|_1$$

l'algorithme “Iterative Shrinkage-Thresholding Algorithm (ISTA)” utilise la récursion

$$\mathbf{x}_{k+1} = \mathcal{T}_{\lambda t}(\mathbf{x}_k - 2t\mathbf{A}^T(\mathbf{A}\mathbf{x}_k - \mathbf{b}))$$

où t est un pas judicieusement choisi. Pouvez vous rappeler le rôle de la fonction $\mathcal{T}_a(\mathbf{x})$ définie de \mathbb{R}^n dans \mathbb{R}^n ?