

Model-based learning

Examen du 30/01/2017

Durée : 2h - Tous documents autorisés

1 Modèles de mélanges dans \mathbb{R} (14 points)

Considérons un problème de classification en K classes dans \mathbb{R} utilisant un modèle de mélange gaussien. Soit un échantillon $(X_1, Z_1), \dots, (X_n, Z_n)$, où X_i est la valeur de la variable d'intérêt pour le i ème individu et $Z_i = (Z_{ik})_{1 \leq k \leq K}$ avec $Z_{ik} = 1$ si l'observation i appartient au groupe k et 0 sinon. Chaque classe est modélisée par une distribution gaussienne d'espérance μ_k et de variance σ_k^2 . Les proportions du mélange sont notées π_k .

1. Considérons un modèle \mathcal{M} particulier supposant $\sigma_k^2 = \sigma^2$. Quelles sont les hypothèses sur les variables faites par ce modèle ?
2. En supposant que les espérances et les proportions du mélange sont connues, calculer l'estimateur du maximum de vraisemblance pour le paramètre σ^2 du modèle \mathcal{M} .
3. Si cette fois les espérances et proportions n'étaient pas connues, quel système devrait-on résoudre pour trouver les estimateurs du maximum de vraisemblance pour l'ensemble des paramètres ? Comment le résoudriez-vous ? (On ne vous demande pas de résoudre ce système).
4. Supposons maintenant que tous les paramètres ont été estimés. Comment classer une nouvelle observation x^* par la règle du maximum a posteriori pour le modèle \mathcal{M} .
5. Considérons deux groupes ($K = 2$) paramétrés par $\{\pi_1, \mu_1, \mu_2, \sigma^2\} = \{0.3, 1, -1, 2\}$. Dans quelle classe classez-vous le point $x = -0.25$?
6. Supposons que l'on considère un second modèle : le modèle complet considérant des variances σ_k^2 différentes par classe. Comment comparer ces deux modèles et choisir le meilleur à utiliser en prédiction ?

Considérons cette fois le même modèle mais dans un objectif de clustering. Les Z_i sont donc inconnus.

7. Ecrire la vraisemblance de modèle. En quoi sa maximisation est différente du cas supervisé ? Comment réaliser cette maximisation (on ne vous demande pas de le faire) ?
8. Puisqu'on est cette fois dans un cadre non supervisé, le nombre de clusters K est inconnu. Pouvez-vous l'estimer par maximum de vraisemblance comme les autres paramètres ?

2 Modèles de mélanges dans \mathbb{R}^p (6 points)

Nous considérons l'utilisation de modèles de mélanges gaussiens p -dimensionnel, dans un cadre non supervisé.

1. Dans l'article Celeux & Govaert (1995)¹, des modèles de mélanges parcimonieux sont introduits. Pour quelle(s) raison(s) ? Pourquoi les contraintes sont-elles imposées sur les matrices de variance et pas sur les autres paramètres ?
2. Aurait-on pu définir des modèles avec des contraintes sur les espérances ? Si oui quel(s) type(s) de contraintes pourrai(en)t être intéressant(s) ?
3. Quelles sont les utilités des critères ICL et BIC ? En quoi sont-ils différents ?

1. Celeux, G. and Govaert, G. (1995). Gaussian parcimonious models. Pattern Recognition, 28(5), 781–793.