EXAMEN DE CLASSIFICATION - 3TSI

Lundi 26 novembre 2018

Polycopié de cours et transparents autorisés

Exercice 1 : Analyse en composantes principales (6 points)

Le tableau suivant indique les résultats de $N = 9$ élèves dans $p = 5$ matières que l'on cherche à analyser à l'aide d'analyses en composantes principales (ACP).

Sample	Maths	Sciences	French	Latin	Music
$\boldsymbol{x}_1$	6	6	5	5.5	8
$\boldsymbol{x}_2$	8	8	8	8	9
$\boldsymbol{x}_3$	6	7	11	9.5	11
$\boldsymbol{x}_4$	14.5	14.5	15.5	15	8
$\boldsymbol{x}_5$	14	14	12	12	10
$\boldsymbol{x}_6$	11	10	5.5	7	13
$\boldsymbol{x}_7$	5.5	7	14	11.5	10
$\boldsymbol{x}_8$	13	12.5	8.5	9.5	12
$\boldsymbol{x}_9$	9	9.5	12.5	12	18

1. (1 pt) Quelle est la normalisation à effectuer avant d'effectuer l'ACP des individus ? Expliquer ce que devient la note 6 obtenue par l'individu $\boldsymbol{x}_1$ dans la matière "Maths" après normalisation (sans effectuer les calculs).

2. (1 pt) Les valeurs propres de la matrice de covariance sont $\lambda_1 = 0.0004, \lambda_2 = 0.0039, \lambda_3 = 0.9831, \lambda_4 = 1.1507$ et $\lambda_5 = 2.8618$. Expliquer comment choisir le nombre de composantes principales p à retenir pour l'ACP des individus.

3. (1 pt) Les résultats de l'ACP des individus dans le plan des deux premières composantes principales sont représentés sur la figure 1a. Expliquer la signification des deux premières composantes principales pour cet exemple.

4. (1 pt) La projection du premier individu dans le premier plan principal est $\widetilde{\boldsymbol{x}_1} = (-2.7857, -0.6765)^T$. Déterminer la qualité de la représentation de cet individu dans le plan formé par les deux premières composantes principales.

5. (1 pt) Analyser les corrélations entre les différentes variables à l'aide de la figure 1b.

6. (1 pt) Expliquer à l'aide de quelques exemples comment on peut utiliser la figure 1c.

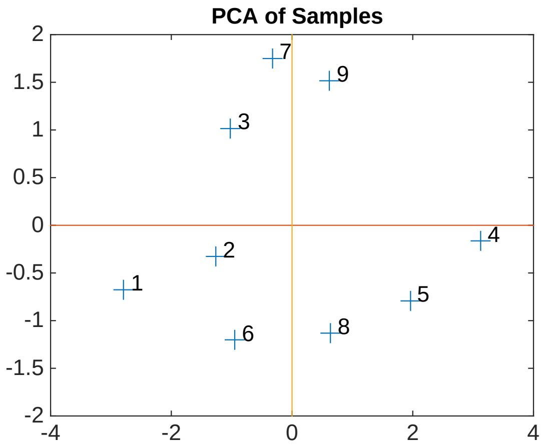
Exercice 2 : Classification Bayésienne (4 points)

On considère un problème de classification à trois classes ω_1, ω_2 and ω_3 de densités normales

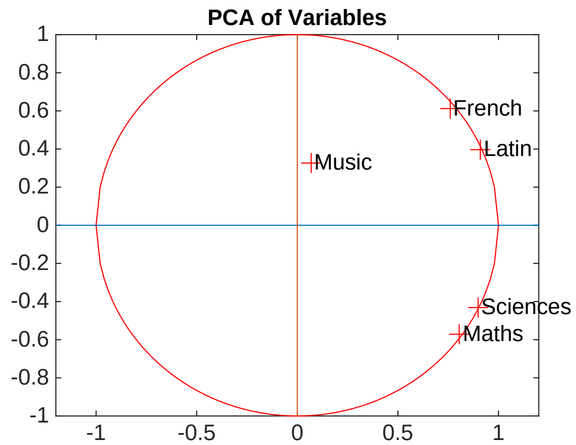
$$f(x|\omega_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{x^2}{2\sigma_i^2}\right), x \in \mathbb{R} \quad i = 1, 2, 3 \quad \text{avec} \quad \sigma_1 > \sigma_2 > \sigma_3. \tag{1}$$

1. (2 pts) Déterminer la règle de classification associée à ce problème avec la fonction de coût $0 - 1$ et lorsque les trois classes sont équiprobables (on admettra que la fonction g telle que $g(u) = \frac{\ln(u)}{u-1}$ est décroissante sur $\mathbb{R}^+$ et on exprimera la règle de classification à l'aide de $T(x) = x^2$).

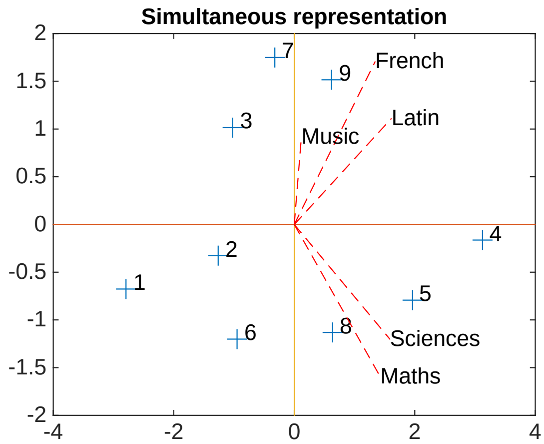
2. (2 pts) On rappelle que si $X \sim \mathcal{N}(0, 1)$ alors $Y = X^2$ suit une loi du chi deux à 1 degré de liberté dont la fonction de répartition est notée ϕ. Expliquer comment calculer la probabilité d'erreur associée à ce classifieur en fonction de $\sigma_1, \sigma_2, \sigma_3$ et de la fonction ϕ (on pourra se limiter au calcul d'un des termes de cette probabilité d'erreur).



(a) PCA of samples.



(b) PCA of variables.



(c) Simultaneous PCA.

Figure 1: Three principal component analyses.

Questions sur l'article (10 **points)**

Remarque : toutes vos réponses doivent être justifiées avec soin.

1. (1 pt) Expliquer le terme "Novelty Detection" utilisé dans le titre.
2. (1 pt) A quoi correspond la courbe d'équation $(\boldsymbol{w}.\Phi(\boldsymbol{x})) - \rho = 0$?
3. (1 pt) Expliquer le role de la variable ξ_i dans (4) et pourquoi le terme $\sum_i \xi_i$ apparait dans la fonction à minimiser.
4. (1 pt) Quel est l'intérêt du terme de régularisation $\frac{1}{2}||w||^2$ intervenant dans (3) ?
5. (1 pt) Quelle est l'expression du Lagrangien associé au problème d'optimisation défini par (3) et (4) ?
6. (1 pt) Expliquer comment on obtient le problème dual défini par (6).
7. (1 pt) Expliquer le rôle du paramètre ν ?
8. (1 pt) Expliquer les résultats de la figure 1.
9. (1 pt) Expliquer comment les auteurs proposent de résoudre un problème de classification (par exemple le problème de reconnaissance de caractères qui donne les résultats de la figure 2).
10. (1pt) A quoi correspondent les caractères représentés sur la figure 2 ?