Examen d'analyse de données - Durée 1h

Les documents de cours (polycopié, TP, TD et notes manuscrites) sont autorisés. Les trois exercices sont indépendants.

1 Classification bayésienne

On considère un problème de classification à deux classes ω_1 et ω_2 pour lequel les données d'entrée $x = (x_1, x_2)^\top$ sont dans $\mathbb{R}^2$. On suppose que les densités de x conditionnellement aux classes ω_1 et ω_2 sont gaussiennes, de moyennes $\mu_1 = (\alpha, \alpha)$ et $\mu_2 = (-\alpha, -\alpha)$, avec $\alpha > 0$, et partagent la même matrice de variance/covariance $\Sigma = \sigma^2 \mathbb{I}_{2\times 2}$, où $0 < \sigma < \alpha$ et $\mathbb{I}_{2\times 2}$ désigne la matrice identité de taille 2×2. Déterminez la règle de décision bayésienne associée à ce problème, pour le coût non informatif 0-1 et dans le cas de deux classes a priori équiprobables. Représentez dans le plan les régions de décision associées à chaque classe. Que se passe-t-il si $0 < \alpha \ll \sigma$?

Pour rappel, la densité de probabilité d'une loi normale en dimension d, notée $N(\mu, \Sigma)$, s'écrit :

$$f(x) = \frac{1}{(2\pi)^{d/2} \, |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right] \tag{1}$$

2 Analyse en composantes principales

On compte les ordres de déplacement *Sauter, Inverser* donnés à un drone par 5 individus qui font effectuer le même parcours à l'engin. On obtient les données suivantes :

Utilisateur	*Sauter*	*Inverser*
Ind. 1	5	7
Ind. 2	3	4
Ind. 3	6	5
Ind. 4	5	5
Ind. 5	6	4

Ces ordres vous semblent-ils corrélés ? Expliquez votre réponse. Représentez les données et les ingrédients résultant de l'ACP de ces données, que vous effectuerez.

3 Modélisation paramétrique de données

On étudie l'absence de stress d'un groupe de 10 étudiants à différents instants, lors d'une épreuve de 3h. On observe que 1h avant l'épreuve ($t = -1$), aucun (0) étudiant n'est décontracté et qu'au début de l'examen, à la lecture du sujet ($t = 0$), 5 étudiants se sentent à l'aise alors qu'après 1h d'examen ($t = 1$), il y a 4 étudiants décontractés. Au bout de 2h d'épreuve ($t = 2$), seuls 3 étudiants restent décontractés. La modélisation suivante est proposée pour exprimer le taux d'absence de stress en fonction du temps t exprimé en heures :

$$g(t) = a\,t^3 + b\,t^2 + c\,t + d \tag{2}$$

où (a, b, c, d) sont des coefficients réels. En posant $\beta = (a, b, c, d)^\top$, écrivez matriciellement le problème aux moindres carrés linéaires à résoudre, c'est-à-dire définissez $A \in \mathbb{R}^{4\times 4}$ et $B \in \mathbb{R}^4$ tels que $\widehat{\beta}_{OLS}$ soit la solution du problème suivant :

$$\min_{\beta \in \mathbb{R}^4} \|A\beta - B\|^2$$