



## Partiel Analyse de Données

Documents autorisés : planches de cours, sujets de TD/TP, notes MANUSCRITES PERSONNELLES de cours/TD (PAS de PHOTOCOPIES), pas de calculatrice.  
Durée : 1h30 (+30 min tiers temps)

### 1 Questions de cours

1. Qu'appelle-t-on méthode de classification supervisée ?
2. Expliquer le principe de la validation croisée.
3. Expliquer ce que sont les vecteurs supports dans la méthode de classification SVM (support vector machines).
4. Dans quelle situation est-il intéressant d'utiliser un noyau dans la méthode de classification SVM ?
5. Donner l'expression mathématique d'une fonction sigmoïde dérivable en tout point utilisée dans les réseaux de neurones.

### 2 Barbecue !

Avec l'arrivée des beaux jours, le barbecue est un repas très apprécié en soirée. Nous essayons dans ce problème de modéliser les ventes de brochettes sur deux mois entre le 30 avril et le 30 juin dans un petit supermarché.

Tout d'abord, on constate qu'au 30 avril ( $t = 0$ ), 10 barquettes de brochettes ont été vendues. Au 30 mai ( $t = 1$ ), seule 1 barquette a été vendue. On décide de modéliser les ventes par la fonction  $f$  suivante :

$$f(t) = a_0 t^2 + b_0$$

avec  $(a_0, b_0)$  des réels et  $t$  exprimé en mois.

1. Résoudre le système linéaire permettant de satisfaire les ventes observées.
2. On remarque qu'au 30 juin ( $t = 2$ ), 16 barquettes de brochettes ont été vendues. Calculer l'erreur aux moindres carrés réalisée par cette modélisation.

Comme ce modèle n'est pas optimal, on décide d'utiliser une autre fonction  $g$  pour modéliser les ventes :

$$g(t) = a t^3 + b t + c \tag{1}$$

où  $(a, b, c)$  sont des coefficients réels.

3. En posant  $\beta = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ , écrivez matriciellement le problème d'estimation aux moindres carrés à résoudre à partir des données de l'énoncé à  $t = \{0, 1, 2\}$ , c'est-à-dire définissez  $A \in \mathbb{R}^{3 \times 3}$  et  $B \in \mathbb{R}^3$  tels que le problème aux moindres carrés puisse s'écrire :

$$\min_{\beta \in \mathbb{R}^3} \frac{1}{2} \|A\beta - B\|^2 \quad (2)$$

4. Dans le cas général d'un problème aux moindres carrés où la matrice  $A \in \mathbb{R}^{m \times n}$ , avec  $m > n$ , donnez la solution théorique du problème (2) sans la calculer explicitement.

### 3 Après le barbecue, un peu de cardio

La Fréquence Cardiaque Maximum, notée FCM, est un paramètre essentiel pour permettre au coureur de fond d'élaborer des plans d'entraînement efficaces. Cette fréquence peut se mesurer, soit en laboratoire sur tapis roulant, soit sur le terrain à l'aide d'un cardio-fréquencemètre. Une première étude a été faite auprès de 5 hommes s'entraînant régulièrement (2 à 4 fois par semaine), et participant à de petites compétitions. On a mesuré leur fréquence cardiaque maximum. On souhaite étudier une relation éventuelle entre l'âge d'un individu et sa fréquence cardiaque maximum.

	Age	FCM
Homme 1	26	178
Homme 2	28	176
Homme 3	30	182
Homme 4	32	180
Homme 5	34	184

1. Calculer la matrice de variance-covariance  $\Sigma$ . Y a-t-il une dépendance entre l'âge et la fréquence cardiaque maximum ? Expliquer votre réponse.
2. Calculer les coordonnées du premier axe principal de ce jeu de données.
3. Représenter sur un même graphe les données, l'axe principal et les composantes principales.
4. Pouvez vous prédire la fréquence cardiaque maximum d'une personne de 38 ans ? Expliquer votre démarche et donner une valeur.