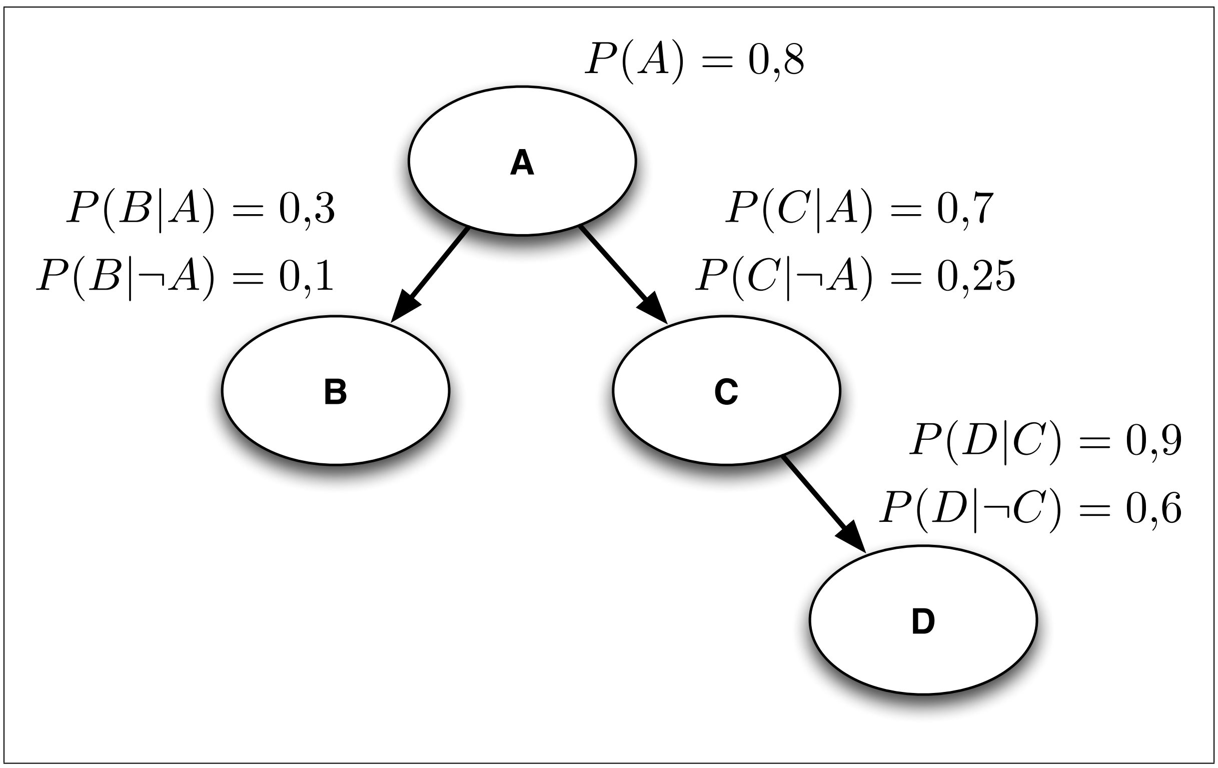
EXAMEN PARTIEL

Instructions : – Une feuille aide-mémoire recto verso manuscrite est permise ;
– Durée de l'examen : 2 h 50.

Pondération : Cet examen compte pour 35% de la note finale.

Question 1 (10 points sur 100)

Soit le réseau bayésien suivant.



(5) (a) Selon ce réseau, calculez la valeur de la probabilité $P(B|\neg C)$.

(5) (b) Toujours selon ce réseau, calculez la valeur de la probabilité $P(D|B)$.

Question 2 (18 points sur 100)

Supposons que l'on fait du classement paramétrique selon deux classes et une variable en entrée (x scalaire), en modélisant les données de chaque classe par une loi normale :

$$p(x|C_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu_i)^2}{2\sigma^2}\right], i = 1{,}2.$$

La variance de chaque classe est la même, $\sigma_1 = \sigma_2 = \sigma$, alors que les moyennes μ_i et les probabilités a priori $P(C_i)$ sont différentes pour chaque classe. Sans perte de généralité, vous pouvez supposer que la moyenne de la classe 1 est inférieure à la classe 2, $\mu_1 < \mu_2$.

(12) (a) Avec cette modélisation, donnez l'équation analytique décrivant les frontières de décision entre les deux classes, en supposant une fonction de perte zéro-un (valeur égale pour les deux types d'erreurs). En une dimension, de telles frontières se résument à des seuils sur la valeur de x. Indiquez également de quelle façon le classement se fait dans les différentes régions séparées par les frontières.

(6) (b) Supposons maintenant que l'on utilise une fonction avec une perte variable selon le type d'erreur, en utilisant une perte $\lambda_{1,2}$ lorsque l'on assigne une donnée de la classe C_1 alors qu'elle appartient à la classe C_2, et inversement en utilisant une perte $\lambda_{2,1}$ lorsqu'une donnée est assignée à C_2 alors qu'elle appartient à C_1. Déterminez la frontière de décision correspondant à ce cas.

Question 3 (12 points sur 100)

Soit le jeu de données des Iris de Fisher, comprenant 150 données en 4 dimensions et organisées selon trois classes. Les estimations du vecteur moyen $\mathbf{m}$ et de la matrice de covariance $\mathbf{S}$ pour l'ensemble des données de ce jeu, en ne tenant pas compte des étiquettes de classes, sont les suivantes :

$$\mathbf{m} = \begin{bmatrix} 0{,}6093 \\ -0{,}0727 \\ 1{,}3404 \\ 0{,}5694 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 0{,}6857 & -0{,}0393 & 1{,}2737 & 0{,}5169 \\ -0{,}0393 & 0{,}1880 & -0{,}3217 & -0{,}1180 \\ 1{,}2737 & -0{,}3217 & 3{,}1132 & 1{,}2964 \\ 0{,}5169 & -0{,}1180 & 1{,}2964 & 0{,}5824 \end{bmatrix}.$$

De plus, les vecteurs propres $\mathbf{c}_i$ et valeurs propres λ_i associées extraits de la matrice de covariance $\mathbf{S}$ sont :

$$\mathbf{c}_1 = \begin{bmatrix} 0{,}6565 \\ 0{,}7297 \\ -0{,}1758 \\ -0{,}0747 \end{bmatrix}, \quad \mathbf{c}_2 = \begin{bmatrix} 0{,}5810 \\ -0{,}5964 \\ -0{,}0725 \\ -0{,}5491 \end{bmatrix}, \quad \mathbf{c}_3 = \begin{bmatrix} 0{,}3616 \\ -0{,}0823 \\ 0{,}8566 \\ 0{,}3588 \end{bmatrix}, \quad \mathbf{c}_4 = \begin{bmatrix} -0{,}3173 \\ 0{,}3241 \\ 0{,}4797 \\ -0{,}7511 \end{bmatrix},$$

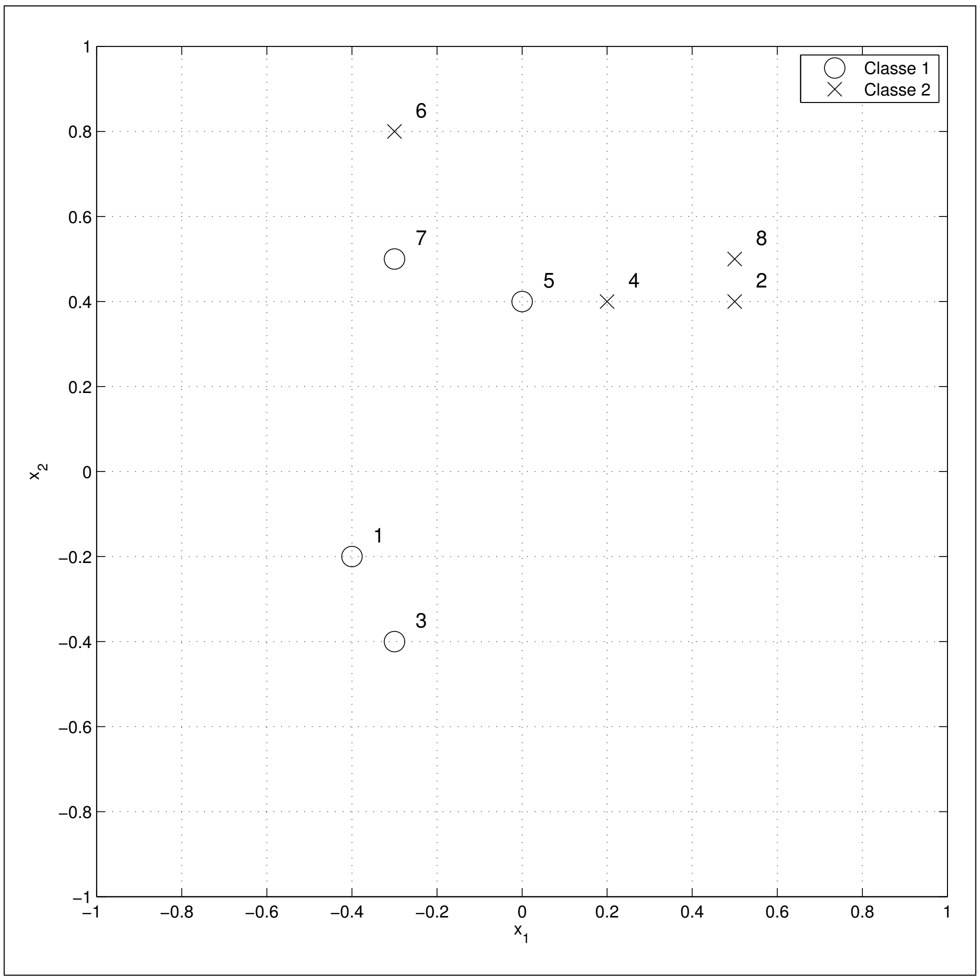
$$\lambda_1 = 0{,}2422, \quad \lambda_2 = 0{,}0785, \quad \lambda_3 = 4{,}2248, \quad \lambda_4 = 0{,}0237.$$

(6) (a) Donner la transformation linéaire $\mathbf{z} = \mathrm{f}(\mathbf{x})$ permettant de projeter les données dans un espace à **deux** dimensions, tout en maximisant la variance conservée. Pour ce faire, donnez l'équation matricielle correspondant à la transformation linéaire et détaillez les valeurs et la signification de chacune des variables de cette équation.

(6) (b) Tracez la courbe de contour correspondant à une distance de Mahanalobis de 1 de la distribution des données dans l'espace à deux dimensions de la sous-question précédente (l'espace des $\mathbf{z}$). Annotez votre tracé en y ajoutant toute l'information nécessaire pour préciser la position, taille et orientation de la courbe de contour.

Question 4 (12 points sur 100)

Soit les données suivantes, en deux dimensions.



(6) (a) Calculez le taux d'erreur de classement selon une approche *leave-one-out* avec un classifieur de type k-plus proches voisins, en employant $k = 3$ voisins et la distance de Manhattan. Explicitez la démarche menant au calcul du taux d'erreur.

(6) (b) Effectuez une édition de Wilson de ce jeu de données, en utilisant un voisin ($k = 1$) et une distance euclidienne. Traitez les données dans leur ordre d'indice, c'est-à-dire dans l'ordre x^1,x^2,x^3, . . . ,x^8. Explicitez votre démarche et rapportez les données formant l'ensemble des prototypes après l'édition.

Question 5 (18 points sur 100)

Supposons que l'on veut appliquer l'algorithme Espérance-Maximisation (EM) à un jeu de données à D dimensions, où chaque groupe $\mathcal{G}_i$ est décrit par une loi normale $\mathcal{N}_D(\boldsymbol{\mu}_i,\mathbf{I})$, dont la covariance correspond à la matrice identité :

$$p(\mathbf{x}|\boldsymbol{\mu}_i) = \frac{1}{(2\pi)^{0,5D}} \exp\left[-\frac{\sum_j (x_j - \mu_{i,j})^2}{2}\right].$$

De plus, les probabilités a priori sont égales pour les K groupes, $P(\mathcal{G}_i) = 1/K$, $i = 1, \ldots, K$.

(8) (a) Donnez le développement complet de l'équation permettant l'estimation $\mathbf{m}_i$ du vecteur moyen $\boldsymbol{\mu}_i$ selon un maximum de vraisemblance.

(5) (b) Donnez les équations décrivant les calculs effectués aux étapes E et M de l'algorithme correspondant à cette modélisation.

(5) (c) Quelle modification supplémentaire à l'algorithme EM basé sur cette modélisation reviendrait à l'algorithme K-means présenté en classe ?

Question 6 (30 points sur 100)

Répondez aussi brièvement et clairement que possible aux questions suivantes.

(3) (a) Expliquez en quoi consiste une option de rejet lorsque l'on fait du classement bayésien.

(3) (b) Expliquez quel résultat on devrait obtenir lorsque la variance d'un estimateur d'un paramètre est élevée, selon le compromis biais-variance.

(3) (c) Supposons que l'on fait de la régression linéaire sur des données à D dimensions. Dites combien de paramètres (valeurs scalaires) décrivant le modèle de régression doivent être estimés.

(3) (d) Expliquez en quoi consiste la régularisation lorsque l'on fait l'inférence de modèles en apprentissage supervisé.

(3) (e) Dans PRTools, quelle est l'utilité de la fonction `testc` ?

(3) (f) Lorsqu'on mesure une corrélation nulle entre deux variables, peut-on dire que ces variables sont indépendantes ? Justifiez brièvement votre réponse.

(3) (g) Soit l'algorithme de sélection avant séquentielle pour la sélection de caractéristiques. Lorsque l'on utilise cet algorithme selon une approche enveloppe (*wrapper*), chaque sous-ensemble de caractéristiques considéré requiert l'entraînement d'un classifieur pour en évaluer la performance. Supposons que l'on veut sélectionner 10 caractéristiques parmi les 20 disponibles, combien d'entraînements de classifieurs seront nécessaires avec l'algorithme de sélection avant séquentielle ?

Propriété pour simplifier vos calculs : $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$.

(3) (h) Quelle est l'utilité d'un multiplicateur de Lagrange lorsque l'on optimise une fonction ?

(3) (i) Indiquez ce que représente l'équation suivante, permettant de faire du classement avec des densités-mélanges :

$$p(\mathbf{x}|C_i) = \sum_{j=1}^{k_i} p(\mathbf{x}|\mathcal{G}_{i,j})P(\mathcal{G}_{i,j}).$$

Prenez soin d'expliquer le sens des variables formant l'équation.

(3) (j) Expliquez pourquoi une estimation non paramétrique de distribution par histogramme ne fonctionne pas bien lorsque la dimensionnalité des données est élevée.