Consignes :
- Sont interdits : Documents, calculettes, téléphones, écouteurs, ordinateurs, tablettes.
- Il est interdit de composer avec un crayon.
- Votre feuille double d'examen doit porter, à l'emplacement réservé, vos nom, prénom, et signature.
- Cette zone réservée doit être cachée par collage.
- Vos feuilles intercalaires doivent être toutes numérotées.
- Le barème est donné à titre indicatif.

Exercice 1 *(4 pts)*

1. Présenter en trois lignes le principe d'une analyse prédictive de données et celui d'une analyse descriptive.
2. Présenter en six lignes une méthode d'apprentissage statistique supervisée et une non-supervisée : Pour cela on décrira le principe général en parlant du fondement probabiliste, du paradigme génératif ou discriminatif, de la nature du critère optimisé (dans le cas où il y en a un), de l'existence ou non d'une solution analytique, de sa mise enœuvre algorithmique, de la notion de convergence locale ou globale, etc)

Exercice 2 *(9 pts)* On considère un échantillon indépendant $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ d'individus $\boldsymbol{X}_i$ décrits par d variables binaires ($\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})^T \in \{0,1\}^d$) indépendantes. On suppose que ces individus sont issus d'une population hétérogène à K classes inconnues. Soit $Y_i \in [\![1, K]\!]$ la classe inconnue de l'individu $\boldsymbol{X}_i$. On dispose d'un échantillon observé $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ où $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{id})^T \in \{0,1\}^d$ est la réalisation de $\boldsymbol{X}_i$. On suppose que les individus $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ sont identiquement distribués selon la loi mélange de lois de Bernoulli multivariées définie par :

$$\mathbb{P}(\boldsymbol{X}_i = \boldsymbol{x}_i; \boldsymbol{\theta}) \quad = \quad \sum_{k=1}^{K} \pi_k \, \mathcal{B}(\boldsymbol{x}_i; \boldsymbol{p}_k) \tag{1}$$

où $\pi_k > 0$ sont les proportions du mélange vérifiant $\sum_{k=1}^{K} \pi_k = 1$, et $\mathcal{B}(\boldsymbol{x}_i; \boldsymbol{p}_k)$ la loi de la $k$ème classe.

1. Justifier pourquoi a-t-on $\mathcal{B}(\boldsymbol{x}_i; \boldsymbol{p}_k) = \prod_{j=1}^{d} \mathcal{B}(x_{ij}; p_{kj})$ où $\mathcal{B}(x_{ij}; p_{kj})$ est la loi de Bernoulli univariée de paramètre $p_{kj} \in]0,1[$ associée à la variable x_{ij} et définie par $\mathcal{B}(x; p) = p^x (1-p)^{1-x}$.
2. Définir le vecteur paramètre $\boldsymbol{\theta}$ et donner le nombre de paramètres libres $\nu_{\boldsymbol{\theta}}$ du modèle
3. On note par $L(\boldsymbol{\theta})$ (respectivement $\log L(\boldsymbol{\theta})$) la vraisemblance (respectivement log-vraisemblance) de $\boldsymbol{\theta}$ pour les données observées $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$. Montrer que la maximisation de la log-vraisemblance $\log L(\boldsymbol{\theta})$ par l'algorithme EM conduit aux étapes suivantes à chaque itération (t) :

   I.Étape E : Cette étape nécessite seulement le calcul des probabilités suivantes :

   $$\tau_{ik}^{(t)} = \frac{\pi_k^{(t)} \, \mathcal{B}(\boldsymbol{x}_i; \boldsymbol{p}_k^{(t)})}{\sum_{\ell=1}^{K} \pi_\ell^{(t)} \, \mathcal{B}(\boldsymbol{x}_i; \boldsymbol{p}_\ell^{(t)})}$$

   II.Étape M : La mise à jour des paramètres du modèle $\boldsymbol{\theta}^{(t+1)}$ s'effectue selon les formules suivantes :

   $$\pi_k^{(t+1)} = \frac{\sum_{i=1}^{n} \tau_{ik}^{(t)}}{n}$$

   et

   $$\boldsymbol{p}_k^{(t+1)} = \frac{\sum_{i=1}^{n} \tau_{ik}^{(t)} \boldsymbol{x}_i}{\sum_{i=1}^{n} \tau_{ik}^{(t)}}.$$

4. Après la convergence de l'algorithme, on obtient le vecteur paramètre estimé du modèle $\widehat{\boldsymbol{\theta}}$. En déduire l'expression d'une règle pour déterminer une estimation de la classe $\widehat{y}_i$ d'un individu $\boldsymbol{x}_i$.
5. On dispose des modèles estimés décrits par $\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_{Kmax}$ où $\widehat{\boldsymbol{\theta}}_K$, $K \in [\![1, Kmax]\!]$, correspond à l'estimation par maximum de vraisemblance du vecteur paramètre d'un modèle de la forme de (1). Donner l'expression d'un critère pour sélectionner le nombre optimal des classes $\widehat{K}$.

Exercice 3 *(7 pts)*

0.5 pt si réponse correcte, -0.5 pt si réponse incorrecte, 0 en cas de non réponse.

Une seule réponse est correcte.

Feuille à remettre. *NE PAS INSCRIRE VOS NOM ET PRENOM SUR CETTE FEUILLE*

Questions	**Réponses**
1. K-nn (K-ppv) est une méthode de classification par apprentissage	☐ Oui ☐ Non
2. K-nn (K-ppv) est un classifieur qui passe facilement à l'échelle	☐ Oui ☐ Non
3. La régression logistique est une méthode	☐ supervisée ☐ non supervisée
4.	☐ générative ☐ discriminative
5.	☐ qui admet une solution analytique ☐ qui nécessite un algorithme d'optimisation
6. L'analyse linéaire discriminante (LDA) est une méthode	☐ supervisée ☐ non supervisée
7.	☐ générative ☐ discriminative
8. En LDA, l'estimation des paramètres	☐ nécessite un algorithme d'optimisation ☐ s'effectue de façon exacte
9. En LDA, la vraisemblance maximisée est	☐ concave ☐ non-concave
10. LDA suppose une matrice de covariance	☐ différente pour toutes les classes ☐ commune à toutes les classes
11. K-means est un algorithme qui se base sur un formalisme probabiliste	☐ Oui ☐ Non
12. K-means est insensible à l'ordre de présentation des données	☐ Oui ☐ Non
13. Dans le cas de son utilisation pour un mélange Gaussien (GMM), EM est insensible à l'ordre de présentation des données	☐ Oui ☐ Non
14. EM garantit l'obtention d'un	☐ optimum global ☐ optimum local