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 *(8 points)*

On considère un échantillon indépendant $((\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_n, Y_n))$ d'individus $\boldsymbol{X}_i$ décrits par p prédicteurs réels ($\boldsymbol{X}_i \in \mathbb{R}^p$) d'une population à 2 classes telle que $Y_i \in [\![0, 1]\!]$ est la classe de l'individu $\boldsymbol{X}_i$. On dispose d'un échantillon observé $((\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n))$ issu de cette même population. On s'intéresse à prédire les classes pour un échantillon de test issu de cette même population à travers le modèle probabiliste défini par :

$$\mathbb{P}(Y_i = 1|\boldsymbol{X}_i = \boldsymbol{x}_i; \boldsymbol{\theta}) = \frac{\exp(\beta_0 + \boldsymbol{\beta}^\top \boldsymbol{x}_i)}{1 + \exp(\beta_0 + \boldsymbol{\beta}^\top \boldsymbol{x}_i)}, \tag{1}$$

où $\boldsymbol{\theta} = (\beta_0, \boldsymbol{\beta}^\top)^\top$.

1. Rappeler le nom de ce modèle.
2. S'agit-il d'un modèle génératif ou discriminatif ? Expliquer pourquoi.
3. On cherche à estimer le vecteur paramètre $\boldsymbol{\theta}$ en maximisant la vraisemblance conditionnelle régularisée suivante :

$$L_R(\boldsymbol{\theta}) = \log \mathbb{P}\left(Y_1 = y_1, \ldots, Y_n = y_n | \boldsymbol{X}_1 = \boldsymbol{x}_1, \ldots, \boldsymbol{X}_n = \boldsymbol{x}_n; \boldsymbol{\theta}\right) - \lambda\, \boldsymbol{\theta}^\top \boldsymbol{\theta}. \tag{2}$$

4. Dériver l'algorithme de Newton-Raphson pour maximiser cette log-vraisemblance.
5. En déduire une forme vectorielle de l'équation de mise à jour de l'algorithme.

Exercice 2 *(8 pts)* On considère un échantillon indépendant $((\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_n, Y_n))$ du couple $(\boldsymbol{X}, Y)$ ou $\boldsymbol{X}_i \in \mathbb{R}^p$ et $Y_i \in \mathbb{R}$. On suppose que l'échantillon est issue d'une population hétérogène à K classes inconnues. Soit $Z_i \in [\![1, K]\!]$ la classe inconnue de l'individu $(\boldsymbol{X}_i, Y_i)$. On dispose d'un échantillon observé $((\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n))$ issu de cette même population. On suppose que condionnellement à la classe $Z_i = k$, les données obéissent au modèle suivant :

$$Y_i = \boldsymbol{\beta}_k^T \boldsymbol{X}_i + \sigma_k E_i, \quad \forall i = 1, \ldots, n. \tag{3}$$

où les variables E_i sont i.i.d de densité $\mathcal{N}(0, 1)$. Dans tout cet exercice on suppose que $\boldsymbol{X}$ est déterministe. On suppose également que $\boldsymbol{X}$ et Z sont indépendants.

1. A partir de la modélisation (3), en déduire que la densité conditionnelle de la classe k est définie par :

$$f(y_i|\boldsymbol{X}_i = \boldsymbol{x}_i, Z_i = k; \boldsymbol{\beta}_k, \sigma_k^2) = \mathcal{N}(y_i; \boldsymbol{\beta}_k^T \boldsymbol{x}_i, \sigma_k^2). \tag{4}$$

2. En déduire que l'on a la densité suivante pour le modèle des observations

$$f(y_i|\boldsymbol{X}_i = \boldsymbol{x}_i;\boldsymbol{\theta}) \quad = \quad \sum_{k=1}^{K} \pi_k \,\mathcal{N}(y_i;\boldsymbol{\beta}_k^T\boldsymbol{x}_i,\sigma_k^2), \tag{5}$$

où $\pi_k = \mathbb{P}(Z_i = k;\boldsymbol{\theta})$.

1. Définir le vecteur paramètre $\boldsymbol{\theta}$ et donner le nombre de paramètres libres $\nu_{\boldsymbol{\theta}}$ du modèle
2. Calculer l'estimateur de $\boldsymbol{\theta}$ défini par

$$\widehat{\boldsymbol{\theta}} \in \arg\max_{\boldsymbol{\theta}\in\mathbb{R}^{\nu_{\boldsymbol{\theta}}}} \ln L(\boldsymbol{\theta}), \tag{6}$$

où

$$\ln L(\boldsymbol{\theta}) = \ln f(y_1,\ldots,y_n|\boldsymbol{X}_1 = \boldsymbol{x}_1,\ldots,\boldsymbol{X}_n = \boldsymbol{x}_n;\boldsymbol{\theta}), \tag{7}$$

et donnez ensuite une écriture matricielle de votre.vos résultat.s. On rappelle que la densité $\mathcal{N}(x;\mu,\sigma^2)$ d'une variable réelle X est définie par :

$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \forall x \in \mathbb{R}$$

avec $\mu \in \mathbb{R}$ et $\sigma > 0$.

3. Soient l'ensemble des vecteurs paramètres $\{\widehat{\boldsymbol{\theta}}_1,\ldots,\widehat{\boldsymbol{\theta}}_G\}$ tel que $\widehat{\boldsymbol{\theta}}_K$, $\forall K \in [\![1,G]\!]$, est une solution de (6). Proposer une méthode de sélection de K parmi celles vues en cours et dériver là en définissant la valeur optimale de K.

Exercice 3 *(4 pts)* Présenter, de façon claire, précise, et concise, et en cinq lignes maximum, chacune des méthodes suivantes.

1. HMM
2. Cartes auto-organisées
3. GTM
4. Préciserez la ou les différence.s entre les deux dernières.