Exercice 1 On considère un échantillon indépendant $((X_1, Y_1), \ldots, (X_n, Y_n))$ d'individus décrits par une variables aléatoire réelle $X_i \in \mathbb{R}$ d'une population à 2 classes telle que $Y_i \in [\![0, 1]\!]$ est la classe de l'individu X_i. On dispose d'un échantillon d'apprentissage $((x_1, y_1), \ldots, (x_n, y_n))$ et d'un échantillon de test $((x_{n+1}, y_{n+1}), \ldots, (x_m, y_m))$ issus de cette même population. On s'intéresse à prédire les classes des données des test sur la base d'un modèle probabiliste appris sur les données d'apprentissage.
On considère la régression logistique. La prédiction s'effectue par la règle du maximum a posteriori (MAP) qui consiste à affecter l'individu x à la classe y maximisant la probabilité a posteriori :

$$\widehat{y} = \arg \max_{k \in [\![0,1]\!]} \mathbb{P}(Y = k|X = x; \boldsymbol{\theta}) \tag{1}$$

où

$$\pi(\boldsymbol{x}; \boldsymbol{\theta}) = \mathbb{P}(Y = 1|X = x; \boldsymbol{\theta}) = \frac{\exp(\beta_0 + \beta x)}{1 + \exp(\beta_0 + \beta x)} \tag{2}$$

et $\mathbb{P}(Y = 0|X = x; \boldsymbol{\theta}) = 1 - \pi(x; \boldsymbol{\theta})$ avec $\boldsymbol{\theta} = (\beta_0, \beta)^T$ étant le vecteur paramètre du modèle.
On suppose que $\boldsymbol{\theta} = (1, 1)^T$.

1. Monter que $Y = 1$ si et seulement si $\widetilde{\boldsymbol{x}}^\top \boldsymbol{\theta} \geq 0$ et 0 sinon, où $\widetilde{\boldsymbol{x}} = (1, x)^T$
2. Prédire les classes des observations suivantes :

$$x_1 = -2, \quad x_2 = 1, \quad x_3 = -1, \quad x_4 = 2$$

3. Tracer la frontière de décision entre les deux classes

On suppose que les paramètres du modèle sont inconnus et on cherche à les estimer à partir de l'échantillon d'apprentissage par maximum de vraisemblance.

1. La solution recherchée est-elle unique ? Expliquer pourquoi.
2. Définir les fonctions de vraisemblance et de log-vraisemblance

$$\ln L(\boldsymbol{\theta}) = \ln f(y_1, \ldots, y_n|X_1 = x_1, \ldots, X_n = x_n; \boldsymbol{\theta})$$

3. Calculer l'estimateur du maximum de vraisemblance $\widehat{\boldsymbol{\theta}}$. Peut-on obtenir une solution analytique ?
4. Proposer un algorithme pour maximiser la (log)-vraisemblance
5. Proposer une écriture matricielle de l'expression de mise à jour
6. Discuter le lien avec l'estimateur des moindres carrés ordinaires (MCO)

Exercice 2 On considère un échantillon indépendant $((\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_n, Y_n))$ d'individus $\boldsymbol{X}_i$ décrits par p variables réelles ($\boldsymbol{X}_i \in \mathbb{R}^p$) d'une population de 2 classes telle que $Y_i \in [\![0, 1]\!]$ est la classe de l'individu $\boldsymbol{X}_i$. On dispose d'un échantillon d'apprentissage $((\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n))$ et d'un échantillon de test $((\boldsymbol{x}_{n+1}, y_{n+1}), \ldots, (\boldsymbol{x}_m, y_m))$ issus de cette même population. On s'intéresse à prédire les classes des données des test sur la base d'un modèle probabiliste appris sur les données d'apprentissage.

On considère la régression logistique. La prédiction s'effectue par la règle du maximum a posteriori (MAP) qui consiste à affecter l'individu $\boldsymbol{x}$ à la classe y maximisant la probabilité a posteriori :

$$\widehat{y} = \arg \max_{k \in [\![0,1]\!]} \mathbb{P}(Y = k|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\theta}) \tag{3}$$

où

$$\pi(\boldsymbol{x}; \boldsymbol{\theta}) = \mathbb{P}(Y = 1|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\theta}) = \frac{\exp(\beta_0 + \boldsymbol{\beta}^T \boldsymbol{x})}{1 + \exp(\beta_0 + \boldsymbol{\beta}^T \boldsymbol{x})} \tag{4}$$

et $\mathbb{P}(Y = 0|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\theta}) = 1 - \pi(\boldsymbol{x}; \boldsymbol{\theta})$ avec $\boldsymbol{\theta} = (\beta_0, \boldsymbol{\beta}^T)^T$ étant le vecteur paramètre du modèle.
On suppose que $\boldsymbol{\theta} = (0, 1, 1)^T$.

1. Monter que $Y = 1$ si et seulement si $\widetilde{\boldsymbol{x}}^\top \boldsymbol{\theta} \geq 0$ et 0 sinon, où $\widetilde{\boldsymbol{x}} = (1, \boldsymbol{x}^T)^T$
2. Prédire les classes des observations suivantes :

$$\boldsymbol{x}_1 = (-2, 3)^t, \quad \boldsymbol{x}_2 = (1, -1)^t, \quad \boldsymbol{x}_3 = (-2, 2)^t, \quad \boldsymbol{x}_4 = (1, -2)^t$$

3. Tracer la frontière de décision entre les deux classes

Exercice 3 On considère le cadre de l'exercice précédent. On suppose que les paramètres du modèle sont inconnus et on cherche à les estimer à partir de l'échantillon d'apprentissage par maximum de vraisemblance.

1. La solution recherchée est-elle unique ? Expliquer pourquoi.
2. Définir les fonctions de vraisemblance et de log-vraisemblance

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

3. Calculer l'estimateur du maximum de vraisemblance $\widehat{\boldsymbol{\theta}}$. Peut-on obtenir une solution analytique ?
4. Proposer un algorithme pour maximiser la (log)-vraisemblance
5. Proposer une écriture matricielle de l'expression de mise à jour
6. Discuter le lien avec l'estimateur des moindres carrés ordinaires (MCO)