

- Durée : 1h30
- Documents autorisés : cours et TD machine
- La copie du voisin n'est pas un document autorisé

On veut réaliser la classification de données $\mathcal{D} = \{(x_i, y_i) \in \mathbb{R}^2 \times \{1, 0\}\}_{i=1}^N$. Ces données sont stockées dans $X \in \mathbb{R}^{N \times 2}$ et $Y \in \mathbb{R}^N$ et disponibles dans le fichier `datatrain.mat`.

On considère le modèle de régression logistique suivant :

$$P(y_i = 1|x_i) = \frac{\exp(z_i^\top w + b)}{1 + \exp(z_i^\top w + b)}$$

où le vecteur z_i est formé comme suit :

$$z_i^\top = (x_{i1} \quad x_{i2} \quad x_{i1}^2 \quad x_{i2}^2)$$

avec x_{i1} (resp x_{i2}) la première (resp. deuxième) coordonnée du point x_i .

Remarque : en notation matlab $x_{i1} = X(i, 1)$ et $x_{i2} = X(i, 2)$.

1. Écrire une fonction `[w, b] = RegLogClass(X, Y)` qui calcule les paramètres w et b du modèle et qui affiche à chaque itération la valeur du critère optimisé
2. Écrire une fonction `Ŷ = RegLogVal(Xt)` qui prend en entrée une matrice de données quelconque $X_t \in \mathbb{R}^{N_t \times 2}$ et renvoie $\hat{Y}_t \in \mathbb{R}^{N_t}$ la prédiction de la classe d'appartenance des points dans X_t .
3. Sachant que le fichier `datatest.mat` contient $X_t \in \mathbb{R}^{N_t \times 2}$ et $Y_t \in \mathbb{R}^{N_t}$, écrire un script qui :
 - apprend le modèle de régression logistique sur les données X, Y ,
 - teste le modèle trouvé sur X, Y puis sur X_t, Y_t ,
 - calcule les erreurs de classification sur les données d'apprentissage et les données de test