

## Régression (III) - Régression logistique

**Exercice 1 (\*).** **Régression logistique binaire simple.** On reprend les données relatives au naufrage du RMS Titanic, pour lesquelles on souhaite classifier la survie selon l'âge des passagers.

Préciser l'expression du modèle logistique utilisé ainsi que celle de la fonction d'erreur associée, et, pour cette dernière, rappeler les résultats énoncés en cours. Proposer enfin une solution à ce problème de régression en écrivant un script structuré à l'aide de Python (utiliser directement les outils fournis par `scikit-learn`) et donner une représentation graphique faisant apparaître la frontière de décision. **Ne pas oublier de normaliser les données !**

**Exercice 2 (\*\*).** **Régression logistique binaire.** On considère les données du dossier `data_ex2` contenant quatre fichiers `csv` :

- deux fichiers contenant les données d'entraînement (entrées et sorties),
- deux fichiers contenant les données de test (entrées et sorties).

Ces données sont issues du dataset `iris` de `scikit-learn`<sup>1</sup>, simplifiées dans le cadre de cet exercice.

1. Quelle est la taille de l'échantillon d'apprentissage ? de test ? Que contiennent les données d'entrée ? de sortie ?
2. Représenter les données d'entraînement dans le plan (en abscisse la longueur du pétale, et en ordonnée sa largeur), en différenciant celles appartenant à la classe iris Virginica de celles n'y appartenant pas. Vous semble-t-il pertinent d'utiliser un modèle logistique binaire pour classifier les données d'entrée selon ces deux classes ?
3. Expliciter le paramètre  $\theta$  du modèle logistique binaire, que l'on souhaite déterminer.
4. Implémenter une solution à ce problème de régression, puis donner la valeur du paramètre  $\theta$  obtenue.
5. Ajouter la frontière de décision ainsi obtenue à la représentation de la question 2.
6. Utiliser le modèle logistique obtenu pour classifier les données de test, puis calculer le score obtenu par le modèle.

<sup>1</sup>[https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load\\_iris.html#sklearn.datasets.load\\_iris](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_iris.html#sklearn.datasets.load_iris)