

Théorie bayésienne de la décision - L3 Info SD

TP - feuille n°2 - Outils pour la création de modèle et test

Le but de ce TP est de créer deux fonctions :

- une fonction `show(...)` permettant de visualiser les données d'entraînement, pour décider d'une frontière de décision
- une fonction `test(...)` permettant de tester la frontière de décision sur des données de validation

Dans tout le TP, on supposera que les fichiers contiennent des jeux de données à 1 caractéristique réelle, réparties en 2 classes.

1 Création des outils

- 1) Dans un fichier Python que vous pourrez appeler `tools.py`, créez une fonction `show(...)` :
 - prenant en entrée un nom de fichier (données d'entraînement)
 - affichant :
 - ◇ les effectifs de chaque classe
 - ◇ sur un même graphique les histogrammes de répartition des caractéristiques `x` pour chaque classe
- 2) Dans un autre fichier Python :
 - importez la fonction `show(...)` de votre module `tools`
 - testez cette fonction sur les données `tp1_data_train.txt`
- 3) Dans le fichier `tools.py`, créez une fonction `test(...)` :
 - prenant en entrée un nom de fichier (données de validation) et une frontière de décision
 - utilisant une prédiction correspondant à cette frontière de décision
 - retournant le taux d'erreur et la matrice de confusion
- 4) Dans un autre fichier Python, testez votre fonction sur le jeu de données de validation contenu dans le fichier `tp1_data_valid.txt`

2 Utilisation des outils sur de nouveaux jeux de données

Utilisez les fonctions précédentes pour choisir une frontière de décision correspondant aux données d'entraînement et testez-la sur les données de validation, sur les jeux suivants :

- 1) `tp2_data1`
- 2) `tp2_data2`
- 3) `tp2_data3`