

Théorie bayésienne de la décision - L3 Info SD

TP - feuille n°7 - Distance de Mahalanobis généralisée

1 Présentation

Le but de ce TP est d'utiliser la distance de Mahalanobis généralisée sur les jeux de données du TP5.

2 Premier jeu de données

- 1) Ouvrez les données d'entraînement contenues dans le fichier `tp5_data1_train.txt` dans une DataFrame Panda et créez :
 - une matrice `X_train` avec les caractéristiques `x1` et `x2`
 - un vecteur `y_train` avec les classes `y`
- 2) Calculez les paramètres de la distance de Mahalanobis généralisée.
- 3) En déduire une fonction `prediction(...)` utilisant la distance de Mahalanobis généralisée.
- 4) Utilisez la méthode `plot_decision(...)` fournie dans le fichier `utils.py` pour afficher la frontière de décision et les nuages de points sur un même graphique.
- 5) Testez la règle de décision sur les données de validation contenues dans le fichier `tp5_data1_valid.txt` (taux d'erreur et matrice de confusion). On pourra comparer avec les résultats obtenus au TP5.

3 Deuxième jeu de données

Mêmes questions avec `tp5_data2_train.txt` et `tp5_data2_valid.txt`.