

Théorie bayésienne de la décision - L3 Info SD

TP - feuille n°6 - Utilisation de la règle de Bayes sur des gaussiennes isotropes

1 Présentation

Le but de ce TP est d'utiliser la méthode paramétrique bayésienne faisant l'hypothèse que les données sont des gaussiennes isotropes :

- avec 2 caractéristiques réelles `x1` et `x2`
- réparties en 2 modalités (0 ou 1) désignées par la classe `y`

Vous allez réutiliser 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) Sous l'hypothèse de gaussiennes isotropes, calculez les paramètres de l'estimateur de maximum de vraisemblance.
- 3) En déduire une fonction `prediction(...)` appliquant la règle de Bayes et qui :
 - prend en paramètre un vecteur `x` de caractéristiques
 - retourne la classe prédite
- 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`.