

UEOMA504 : Traitement Statistique de l'information
Master - Deuxième année

Travaux Pratiques

Feuille 2

Préliminaires :

On s'intéresse toujours aux individus âgés de plus de 20 ans dans travaillant dans le domaine de l'informatique, l'objectif étant de les classer par sexe. Le TP précédent ayant montré que notre classification par taille donnait une trop grande probabilité d'erreur, nous envisageons d'utiliser une nouvelle caractéristique pour améliorer les performances de notre classifieur. Nous utiliserons maintenant la taille ET le poid pour nos classifications.

Voici les informations dont on dispose :

- Parmi ces informaticiens, une proportion p_1 est de sexe féminin, $p_0=0.6$ de sexe masculin.
- l'hypothèse de sexe masculin sera notée H_0 , celle de sexe féminin sera notée H_1 . Pour un individu donné, la mesure de sa taille et de son poid sera notée $x=[taille, poid]$
- La densité de probabilité de x suit une loi normale bivariée pour chacune des classes.
 - pour les hommes, la moyenne de taille est de 178 cm, la moyenne de poids de 75 Kg. L'écart type sur les tailles est de 10 (cm), l'écart type de poids de 9 (Kg). Le coefficient de corrélation entre les deux est de 0.6

Questions :

- Afficher sur des graphes différents les densités de probabilités $p(x/w_0)$ et $p(x/w_1)$. La fonction permettant de calculer la densité de probabilité d'une loi normale multivariée est `dmnorm`. La fonction permettant de tracer des surfaces est `persp`.
 - pour les femmes, la moyenne de taille est de 165 cm, la moyenne de poids de 65 Kg. L'écart type sur les tailles est de 9 (cm), l'écart type de poids de 10 (Kg). Le coefficient de corrélation entre les deux est de 0.5
- Afficher le graphe du classifieur bayésien optimal pour l'erreur moyenne. Ce classifieur fonctionne-t-il mieux que celui qui ne tenait compte que de la taille ?

Considérons maintenant qu'il est plus grave de mal classifier une femme qu'un homme, on définira les coûts suivants :

$d_i \setminus w_j$	w_0	w_1
d0	0	30
d1	20	0

- Construire le classifieur bayésien optimal pour le risque total moyen associé à ces coûts.