Université de Bretagne-Sud
L3 Mathématiques, parcours Statistique

Analyse discriminante
TD 1

1 – Soit X une variable aléatoire normalement distribuée sur deux classes ω_1, ω_2 de probabilités a priori $p(\omega_1) = 1/3$ et $p(\omega_2) = 2/3$. On suppose que $X|\omega_1 \sim \mathcal{N}(0, 1)$ et $X|\omega_2 \sim \mathcal{N}(d, 1)$, où d est strictement positif. On pourra faire varier d dans les questions suivantes. Représenter sur un même graphique les densités $p(x|\omega_1)$ et $p(x|\omega_2)$.

2 – Représenter sur un même graphique $p(x|\omega_1)p(\omega_1)$ et $p(x|\omega_2)p(\omega_2)$.

3 – Représenter graphiquement le rapport de vraisemblance $p(x|\omega_1)/p(x|\omega_2)$. Représenter à l'aide d'une droite horizontale le seuil correspondant à la règle de Bayes pour discriminer entre ω_1 et ω_2 au vu d'une observation x.

4 – Exprimer la règle de Bayes à l'aide d'une fonction discriminante simple. Visualiser cette règle sur les graphiques obtenus précédemment.

5 – Représenter sur un même graphique $p(\omega_1|x)$ et $p(\omega_2|x)$. Visualiser la règle de Bayes sur le graphique obtenu.

6 – On rappelle que l'erreur de classification associée à la règle de Bayes est donnée par

$$e_B = 1 - \int_{\mathcal{X}} \max_i p(\omega_i)p(x|\omega_i)\, dx$$

où $\mathcal{X}$ désigne l'espace des mesures. Visualiser cette erreur de classification sur les graphiques obtenus précédemment.

7 – On rappelle qu'on peut définir une région de *rejet* associée au seuil t comme

$$R = \{x : \max_i p(\omega_i|x) < 1 - t\}$$

Visualiser une telle région sur les graphiques obtenus précédemment. Visualiser la valeur de t à partir de laquelle l'option de rejet est activée.

8 – Obtenir formellement et visualiser la règle de Neyman-Pearson de discrimination entre la classe positive ω_1 et la classe négative ω_2 au vu de x, au seuil $\alpha = 5\%$. On rappelle que celle-ci est basée sur le rapport de vraisemblance et qu'elle est obtenue en minimisant la probabilité d'erreur de type I (probabilité d'affecter x à ω_2 alors qu'il appartient à ω_1) sous la contrainte que la probabilité d'erreur de type II (probabilité d'affecter x à ω_1 alors qu'il appartient à ω_2) soit égale à α.

9 – Obtenir des expressions formelles de l'erreur de type I et de l'erreur de type II et les utiliser pour représenter graphiquement la courbe ROC associée au test de Neyman-Pearson *pour différentes valeurs de d*. Comment interpréter le résultat obtenu ?

10 – Répondre aux questions précédentes, dans la mesure du possible, pour $X|\omega_1 \sim \mathcal{N}(0, 1)$, $X|\omega_2 \sim 0.6\,\mathcal{N}(1, 1) + 0.4\,\mathcal{N}(-1, 2)$ et $p(\omega_1) = p(\omega_2) = 1/2$.