

HMMA303 Discrimination et scoring

TP1 Règle et classifieur de Bayes

Jean-Michel Marin

Version du 11/09/2020

On fixe les paramètres associés au modèle de mélange de 2 lois gaussiennes que nous allons étudier. On s'intéresse à un problème de classification binaire $Y \in \{0, 1\}$. On fixe $\pi_0 = \mathbb{P}(Y = 0) = 3/10$ et $\pi_1 = \mathbb{P}(Y = 1) = 7/10$. Par ailleurs, on suppose que la variable X conditionnellement à la classe $Y = k$ est distribuée suivant une loi gaussienne en dimension 2

$$X|Y = k \sim \mathcal{N}_2(\mu_k, \Sigma_k)$$

avec

$$\mu_0 = (3, 3), \mu_1 = (5, 5), \Sigma_0 = \begin{bmatrix} 5 & 45/10 \\ 45/10 & 5 \end{bmatrix} \text{ et } \Sigma_1 = \begin{bmatrix} 5 & -45/10 \\ -45/10 & 5 \end{bmatrix}.$$

- 1 Tracer la densité de la loi marginale de X
- 2 Tracer la probabilité de $Y=0$ sachant X
- 3 Tracer les frontières du classifieur de Bayes
- 4 Simuler un échantillon d'apprentissage et un échantillon de test
- 5 Mettre en oeuvre la classification par regression sur ces données et estimer son erreur
- 6 Mettre en oeuvre la méthodes des k-plus-proches-voisins sur ces données et estimer son erreur
- 7 Estimer le taux d'erreur du classifieur de Bayes