SY19 – A19
TP6: Modèles additifs généralisés pour la discrimination

Exercice 1

Soit $((X_1, X_2), Y)$ la distribution jointe de données dans le plan ainsi que leur classe associée ($Y \in \{0, 1\}$).

$$X_1 \mid (Y = 0) = R_1 \cos\theta_1 \qquad X_1 \mid (Y = 1) = R_2 \cos\theta_2$$
$$X_2 \mid (Y = 0) = R_1 \sin\theta_1 \qquad X_2 \mid (Y = 1) = R_2 \sin\theta_2$$

avec $\theta_1, \theta_2 \sim \mathcal{U}(0, 2\pi)$, $R_1 \sim \mathcal{N}(r_1, 1)$, $R_2 \sim \mathcal{N}(r_2, 1)$ et $\Pr(Y = 1) = \Pr(Y = 0) = 1/2$.

1. Compléter la fonction `sample.donut` du fichier `src/dataset.R` qui génère un jeu de données suivant le modèle précédent. Afficher le jeu de données obtenu pour $r_1 = 3$ et $r_2 = 4$. Quelle semble être la frontière de décision optimale ?
2. Est-ce qu'une analyse discriminante linéaire convient pour discriminer les deux classes ?
3. Refaites une analyse discriminante linéaire en rajoutant des descripteurs qui vont aider à la discrimination.
4. Utiliser un modèle linéaire généralisé avec des splines lissées. On utilisera la fonction `gam` de la bibliothèque `mgcv`.
5. Visualiser la frontière de décision avec la fonction `vis.gam`.
6. Retrouver le caractère quadratique des fonctions apprises sur chacun des prédicteurs.
7. Re-tester l'analyse discriminante linéaire et le modèle additif généralisé avec peu de données. Que se passe-t-il ?