

SY19

TP09: Modèles additifs généralisés pour la discrimination

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

$$\begin{aligned} X_1 \mid (Y = 0) &= R_1 \cos \theta_1 & X_1 \mid (Y = 1) &= R_2 \cos \theta_2 \\ X_2 \mid (Y = 0) &= R_1 \sin \theta_1 & X_2 \mid (Y = 1) &= R_2 \sin \theta_2 \end{aligned}$$

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 $\Pr(Y = 1) = \Pr(Y = 0) = 1/2$.

Partie I : génération des données et ADL

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 $r_2 = 5$. 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.

Partie II : analyse discriminante avec modèles de mélange

1. Utiliser l'analyse discriminante avec modèles de mélange gaussien (mixture discriminant analysis). Combien de composantes trouve-t-on dans chaque classe ? Représenter graphiquement les composantes des mélanges pour les deux classes, et les points mal classés.

Partie III : modèle additif généralisé

1. Utiliser un modèle additif généralisé avec des splines lissées.
Afin de pouvoir visualiser facilement entre autres les splines apprises on utilisera la fonction `gam` de la bibliothèque `mgcv` et non la fonction `gam` de la bibliothèque `gam`. Elle fonctionne de manière similaire à la différence près qu'il faut spécifier l'argument `family` dans `gam` et qu'il n'y a pas besoin de spécifier les degrés de liberté `df` dans la formule.
2. Visualiser la frontière de décision avec la fonction `vis.gam` et les arguments `type` et `plot.type`.
3. Retrouver le caractère quadratique des fonctions apprises sur chacun des prédicteurs.
On utilisera la fonction `plot` avec l'argument `select`.
4. Re-tester l'analyse discriminante linéaire et le modèle additif généralisé avec peu de données.
Que se passe-t-il ?