

**TP3 :****• Exercice 1 : (Régression logistique)**

Trente patients ont reçu un certain niveau de dosage d'agent anesthésique pendant 15 minutes. Puis une incision leur est faite. Il est ensuite noté si le patient a bougé ou pas lors de l'incision. Ainsi, pour chaque patient, on dispose :

- du dosage de l'agent anesthésique pendant 15 minutes (variable  $X_1$ ),
- du fait qu'il ait bougé ou pas (variable  $Y$ , avec  $Y = 1$  pour bougé).

On souhaite expliquer  $Y$  en fonction de  $X_1$ .

Les données sont dans le fichier anesthesie.txt que vous trouverez sur Moodle dans le cours Big Data.

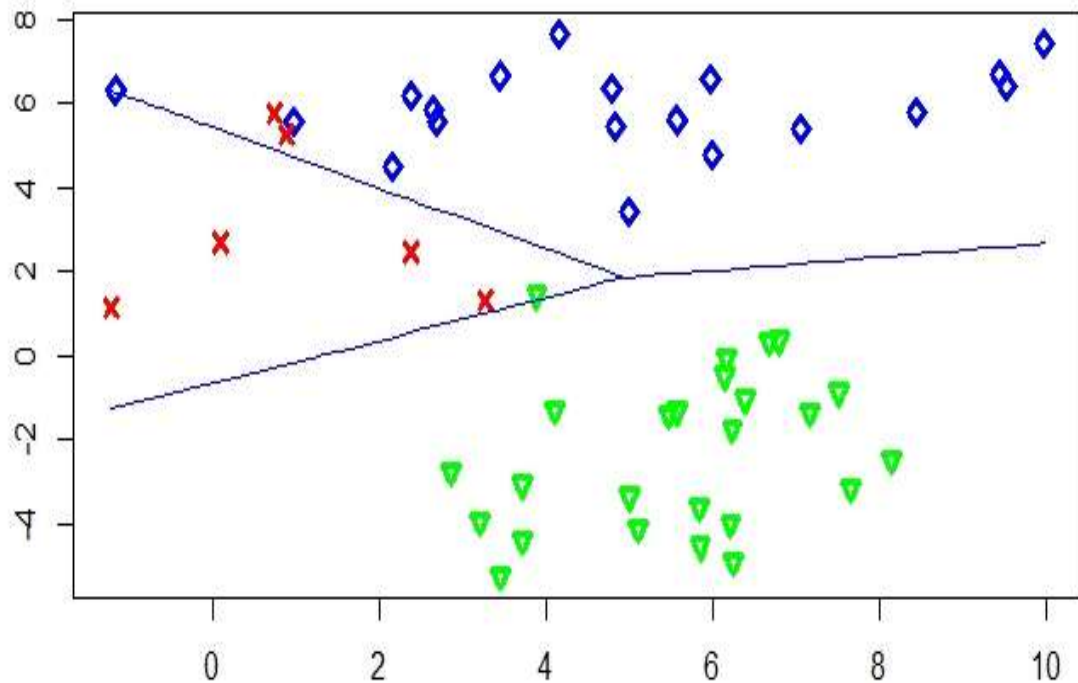
- 1- Modéliser le problème à l'aide d'une régression logistique.
- 2- Est-ce qu'un patient ayant eu pour dosage  $X_1 = 1.25$  a plus de chance de bouger que de ne pas bouger ?
- 3- Faire le graphique des données et de la courbe de régression logistique.
- 4- Calculer un taux d'erreur par validation croisée. Les résultats sont-ils bons ?

**• Exercice 2 : (Analyse en composantes principales)**

Dans cet exercice, nous considérons les statistiques de crimes commis aux États-Unis en 1977. Ce jeu de données contient les nombres de crimes pour 100 000 habitants pour  $p = 7$  types de crimes dans les  $n = 50$  états. Pour récupérer ces données, il faut charger dans R le fichier crimes.txt. Explorer rapidement ces données et calculer l'ACP.

- 1- A l'aide de l'ébouillie des valeurs propres, quel nombre  $d$  de variables principales utiliseriez-vous ?
- 2- A quelle part d'inertie expliquée cela correspond-il ?
- 3- Calculer la contribution des variables initiales à l'inertie des variables principales ?
- 4- Discuter des résultats obtenus par rapport à votre choix de  $d$
- 5- Représenter les variables sur le cercle des corrélations
- 6- A l'aide de ce graphique, comment allez-vous interpréter les axes dans le plan principal ?
- 7- Représenter les individus dans le plan principal avec la fonction plot. Cette représentation est-elle suffisante pour résumer correctement l'information des données ?
- 8- Pour  $d > 2$ , représenter les données dans les plans engendrés par toutes les paires d'axes principaux parmi les  $d$  premiers.
- 9- Calculer la contribution de chaque état à l'inertie des axes.
- 10- Repérer quelques états dont certaines contributions sont importantes.
- 11- Pouvez-vous donner une interprétation géographique aux axes de l'ACP ?

- **Exercice 3 : (Analyse discriminante)**



- 1- Générer 50 données issues de 3 groupes de centres  $(1 ; 2)$ ,  $(6 ; 6)$  et  $(6 ; -2)$  avec probabilité 0.2, 0.3 et 0.5 issues d'une loi gaussienne. (voir exemple ci-dessus)
- 2- Faire une analyse discriminante sur ces données en séparant le plan en trois régions (comme sur le graphique)
- 3- Quel est le taux d'erreur de classification avec cette méthode.
- 4- Proposer une autre méthode et donner son taux d'erreur.