# TP4 Analyse discriminante, méthodes linéaires

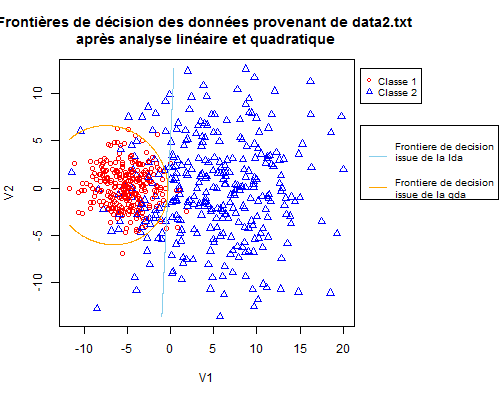
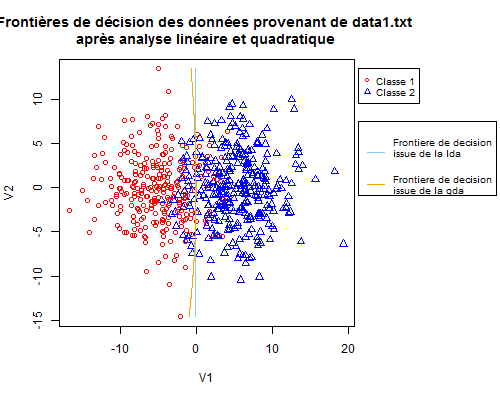
Lors de ce TP, nous étudierons les méthodes d’analyses discriminantes et de régression logistique en déduisant des règles de décisions et en estimant les erreurs de classifications associées, nous comparerons par la suite ces deux outils entre eux. Enfin, nous les utiliserons dans une dernière partie pour établir des règles de décisions quant à la prédiction de diabète sur une population d’amérindiens.

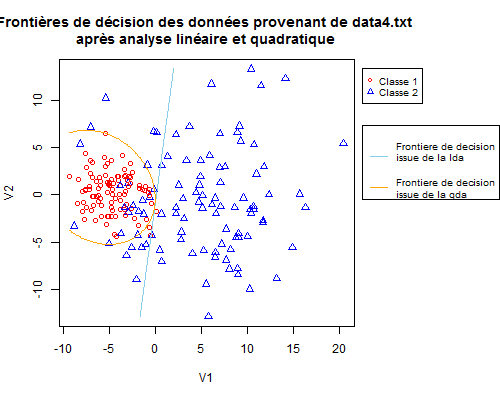
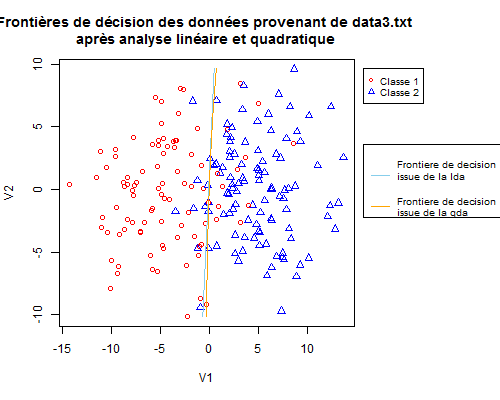
## Exercice 1 Analyse discriminante

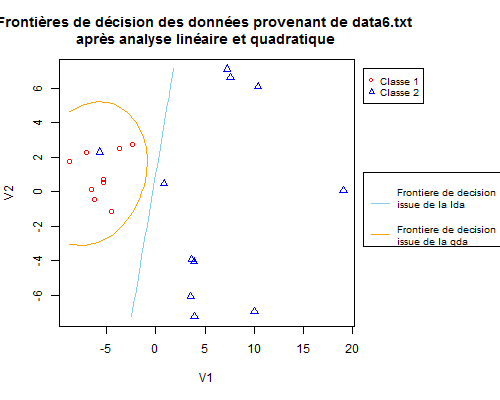
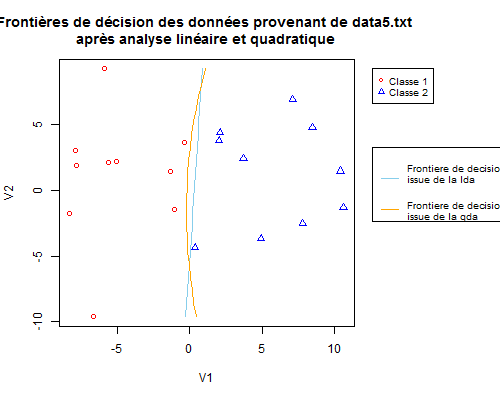
L’analyse discriminante a pour objectif de fournir des frontières de décision grâce à une estimation des paramètres théoriques, faite sur un ensemble d’apprentissage.

(Pour les valeurs d’erreurs de classification voir en ANNEXE)

Question 1







Ci-dessus sont représentés les six jeux de données ainsi que les frontières de décisions respectives de la LDA et de la QDA. En analysant les valeurs des erreurs de classification de chacune de ces analyses discriminantes, on s’aperçoit que la QDA a une erreur de classification moins importante ou égale à celle de la LDA. Nous pouvons interpréter cette observation par l’inflexion de la frontière de décision quadratique (alors que celle de l’analyse discriminante linéaire est toujours droite) qui témoigne d’une meilleure sensibilité du modèle.

### Question 2

Cette fois ci, nous avons appliqué la LDA et la QDA en effectuant l’apprentissage des paramètres sur une partie des jeux de données (ensemble d’apprentissage= 2/3 tirés aléatoirement) uniquement. Nous avons donc pu calculer les erreurs de classifications moyennes pour ces ensembles d’apprentissage mais également sur les ensembles de test (données restantes).

A la lecture des résultats nous observons une erreur systématiquement plus grande pour les ensembles de test. Ceci était prévisible, les paramètres ayant été estimés sans prendre en compte les données contenues dans les ensembles de test. De plus, l’erreur de classification sur les ensembles de test de la QDA observe une plus grande augmentation par rapport aux erreurs sur les ensembles d’apprentissage que la LDA. Cela est également due à la plus grande précision de ce modèle qui dont les décisions collent plus spécifiquement à l’ensemble d’apprentissage.

### Question 3

Nous effectuons la même procédure qu’à la question précédente en faisant varier la proportion de l’ensemble d’apprentissage.

On observe une optimisation de l’erreur

## Exercice 2 Régression logistique

La régression logistique est un modèle de discrimination linéaire permettant de modéliser les probabilités à posteriori par des fonctions linéaires.

### Question 1

Exercice3

ANNEXES

Exercice 1

ex1-data1.txt.png sauvegardee  
data1.txt for p(test) = 1  
        proba erreur lda: 0.09166667   
        proba erreur qda: 0.09166667   
data1.txt for p(test) = 0.25   
        proba erreur lda training: 0.09006667   
        proba erreur qda training: 0.09026667   
        proba erreur lda test: 0.09472222   
        proba erreur qda test: 0.09597778   
data1.txt for p(test) = 0.3333333   
        proba erreur lda training: 0.0909   
        proba erreur qda training: 0.0908   
        proba erreur lda test: 0.094625   
        proba erreur qda test: 0.0956625   
data1.txt for p(test) = 0.5   
        proba erreur lda training: 0.09238333   
        proba erreur qda training: 0.09213333   
        proba erreur lda test: 0.0937   
        proba erreur qda test: 0.09441667   
data1.txt for p(test) = 0.6666667   
        proba erreur lda training: 0.0923   
        proba erreur qda training: 0.09235   
        proba erreur lda test: 0.092825   
        proba erreur qda test: 0.0939   
        variance lda test: 0.0002918536   
        variance qda test: 0.0002879799   
data1.txt for p(test) = 0.75   
        proba erreur lda training: 0.09151111   
        proba erreur qda training: 0.09203333   
        proba erreur lda test: 0.0947   
        proba erreur qda test: 0.09513333   
  
ex1-data2.txt.png sauvegardee  
data2.txt for p(test) = 1  
        proba erreur lda: 0.11   
        proba erreur qda: 0.08333333   
data2.txt for p(test) = 0.25   
        proba erreur lda training: 0.1132333   
        proba erreur qda training: 0.08353333   
        proba erreur lda test: 0.1129556   
        proba erreur qda test: 0.08448889   
data2.txt for p(test) = 0.3333333   
        proba erreur lda training: 0.112875   
        proba erreur qda training: 0.08475   
        proba erreur lda test: 0.1134875   
        proba erreur qda test: 0.0833875   
data2.txt for p(test) = 0.5   
        proba erreur lda training: 0.1114667   
        proba erreur qda training: 0.08246667   
        proba erreur lda test: 0.1115833   
        proba erreur qda test: 0.0836   
data2.txt for p(test) = 0.6666667   
        proba erreur lda training: 0.11035   
        proba erreur qda training: 0.081425   
        proba erreur lda test: 0.11465   
        proba erreur qda test: 0.085425   
        variance lda test: 0.0003189724   
        variance qda test: 0.000207733   
data2.txt for p(test) = 0.75   
        proba erreur lda training: 0.1102   
        proba erreur qda training: 0.08215556   
        proba erreur lda test: 0.1134333   
        proba erreur qda test: 0.08443333   
  
ex1-data3.txt.png sauvegardee  
data3.txt for p(test) = 1  
        proba erreur lda: 0.12   
        proba erreur qda: 0.125   
data3.txt for p(test) = 2/3  
        proba erreur lda training: 0.1218421   
        proba erreur qda training: 0.123609   
        proba erreur lda test: 0.1195522   
        proba erreur qda test: 0.121194   
        variance lda test: 0.0009626882   
        variance qda test: 0.001008742   
  
ex1-data4.txt.png sauvegardee  
data4.txt for p(test) = 1  
        proba erreur lda: 0.13   
        proba erreur qda: 0.095   
data4.txt for p(test) = 2/3  
        proba erreur lda training: 0.1346617   
        proba erreur qda training: 0.08552632   
        proba erreur lda test: 0.1328358   
        proba erreur qda test: 0.09022388   
        variance lda test: 0.001061221   
        variance qda test: 0.0008715834   
  
ex1-data5.txt.png sauvegardee  
data5.txt for p(test) = 1  
        proba erreur lda: 0   
        proba erreur qda: 0   
data5.txt for p(test) = 2/3  
        proba erreur lda training: 0.01038462   
        proba erreur qda training: 0.01576923   
        proba erreur lda test: 0.06428571   
        proba erreur qda test: 0.1378571   
        variance lda test: 0.007127474   
        variance qda test: 0.02058814   
  
ex1-data6.txt.png sauvegardee  
data6.txt for p(test) = 1  
        proba erreur lda: 0.05   
        proba erreur qda: 0.05   
data6.txt for p(test) = 2/3  
        proba erreur lda training: 0.05692308   
        proba erreur qda training: 0.04192308   
        proba erreur lda test: 0.08214286   
        proba erreur qda test: 0.1228571   
        variance lda test: 0.01013999   
        variance qda test: 0.01621167

Exercice 2

for data1.txt :  
        proba erreur reg lineaire 0.093175   
        variance reg lineaire 0.0002578335   
        proba erreur reg quadratique 0.094575   
        variance reg quadratique 0.0002592406   
  
for data2.txt :  
        proba erreur reg lineaire 0.1065   
        variance reg lineaire 0.0003175879   
        proba erreur reg quadratique 0.0833   
        variance reg quadratique 0.0002830251   
  
for data3.txt :  
        proba erreur reg lineaire 0.125597   
        variance reg lineaire 0.001068558   
        proba erreur reg quadratique 0.1223134   
        variance reg quadratique 0.001311296   
  
for data4.txt :  
        proba erreur reg lineaire 0.138806   
        variance reg lineaire 0.001587353   
        proba erreur reg quadratique 0.09895522   
        variance reg quadratique 0.001088781

Exercice 3

erreur lda: 0.2018072   
erreur qda: 0.2289157   
erreur logreg: 0.1987952