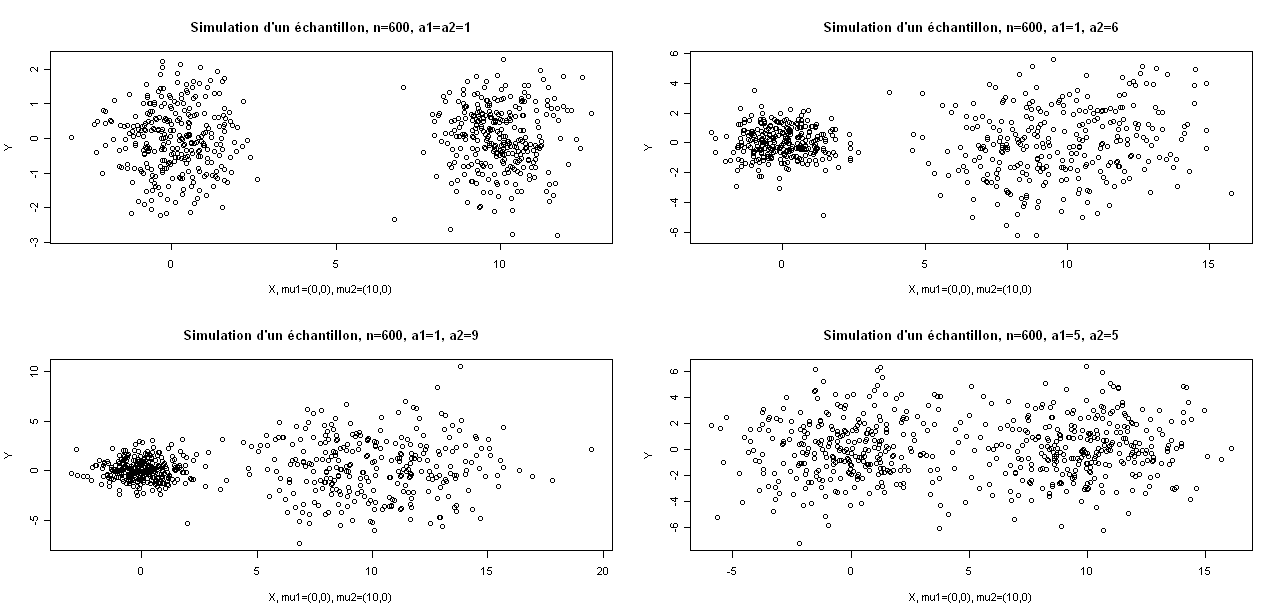
|  |
| --- |
| C:\Users\Thibaud\UTC\logo_UTC.gifSY09 | Rapport TP3  BAGUET Thibaud  HOU Yunhui |
| Théorie de la décision |

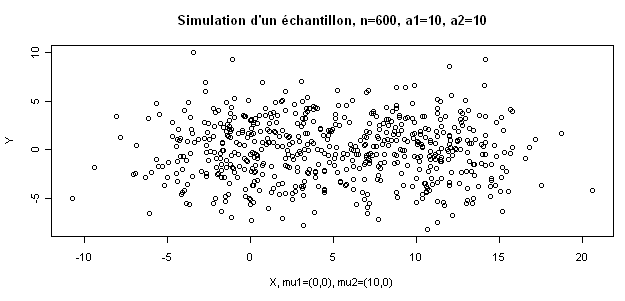
# Objectifs du TP

# Classificateur euclidien

On veut étudier les performances du classificateur euclidien sur des échantillons issus de deux classes *ω*1 et *ω*2 de ***R2*** dont les distributions sont normales et de paramètres (*μ1*, *a1I*) et (*μ2, a2I*).

### Simulation d’un échantillon





|  |
| --- |
| *simul<-function(n,mu1,mu2,a1,a2){*  *sigma1=a1\*diag(2)*  *sigma2=a2\*diag(2)*  *X1=mvrnorm(n/2,mu1,sigma1)*  *X2=mvrnorm(n/2,mu2,sigma2)*  *X1=cbind(X1,matrix(1,nrow=n/2,byrow=T))*  *X2=cbind(X2,matrix(2,nrow=n/2,byrow=T))*  *X=rbind(X1,X2)*  *X=X[sample(1:n),]}* |

On a ici simulé deux échantillons issus d’une loi normale multidimensionnelle avec la fonction mvrnorm de R. Nous pouvons graphiquement le résultat :

* Pour des a faibles (donc des variances faibles), les points sont rapprochés.
* Pour des a élevés (donc une grande variance), les points sont étalés. Ceci est logique car la variance représente le carré de la dispersion des 300 valeurs autour de leur moyenne. (donc plus la variance est grande, plus la dispersion est grande).

### Estimation de la probabilité d’erreur

Maintenant, les moyennes et les variances ne sont pas fixées et nous allons essayer de les retrouver grâce aux fonctions regleEuclidienne et erreurEstimee :

|  |  |
| --- | --- |
| regleEuclidienne<-function(x,mu1,mu2){  cls=2  dis1=sqrt((x[1]-mu1[1])^2+(x[2]-mu1[2])^2)  dis2=sqrt((x[1]-mu2[1])^2+(x[2]-mu2[2])^2)  if (dis1<dis2) cls=1  return(cls)} | erreurEstimee<-function(D,regle,mu1,mu2){  n=0  cls=apply(D,1,regle,mu1=mu1,mu2=mu2)  for(i in 1:dim(D)[1]) {if(cls[i]!=D[i,3]) n=n+1}  n=n/dim(D)[1]  n} |

La fonction regleEuclidienne va retourner les valeurs « 1 » ou « 2 » suivant la classification d’une valeur dans une classe ou dans une autre. En vérifiant, la distance euclidienne séparant une valeur x des centres (µ1 et µ2) des deux classes, la valeur sera affectée à la classe dont le centre est le plus proche. D’autre part, si une valeur x de la classe *w1* est plus éloigné de µ1 que de µ2, alors la fonction lui assimilera la valeur 2, ou inversement.

En utilisant la fonction, erreurEstimee on va calculer la probabilité d’erreur du classificateur euclidien. Ainsi si un échantillon x appartient à la classe *ω*1 et qu’il ressort avec la valeur 2 alors il y aura eu une erreur de classification. Cette fonction va donc permettre de sommer les erreurs et d’en tirer ainsi une probabilité d’erreur estimée.

### Qualité du résultat

Nous avons appliqué la tâche une dizaine de fois et nous obtenons les résultats suivants :

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Essais | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Moy | Var |
| Erreur(%) | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? | ? |

La probabilité moyenne d’erreur du classificateur euclidien est de

# Règles de Neyman-Pearson et de Bayes

### Calcul préliminaire

|  |  |
| --- | --- |
|  |  |

### Estimations des paramètres

MAUVAIS TP : PAS LA MEME QUESTION !!!!!!!!!!!!!!!!!!!!!!!

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

!!!!!!!!!!!!!!!!!!!!!!!!

Nous utilisons donc la fonction *rnorm*. Nous obtenons alors 4 vecteurs : vect1X1 et vect1X2 pour la loi *f1* et vect2X1 et vect2X2 pour la loi *f2*

Nous allons comparer la moyenne des résultats obtenus avec les paramètres des lois.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Taille du jeu | Vect1X1 | | | Vect1X2 | |
| **Moyenne** | **Variance** | | **Moyenne** | **Variance** |
| 300 | -0.9392797 | | 0.849767 | -0.01189009 | 0.9643012 |
| 1000 | -0.9614003 | | 1.038214 | 0.03490163 | 0.9682516 |
| 2000 | -0.9992178 | | 1.041234 | 0.003959581 | 1.032255 |

On remarque que plus il y a de valeurs, plus les valeurs empiriques se rapprochent des valeurs théoriques. On peut en conclure que plus la taille du jeu est important, plus on se rapproche des paramètres de la loi générée. Ceci est explicable grâce à la loi des grands nombres.

### Iso-densité

Chaque variable suit une loi Normale *N* (*µ,σ*) d’équation :

On va donc montrer que et que , avec et constants de manière à obtenir des cercles. On a donc

|  |  |
| --- | --- |
|  |  |

Et donc

|  |  |
| --- | --- |
|  |  |

Or l’équation d’un cercle est donnée par . D’où le résultat suivant :

**et**

### Neyman-Pearson

### Une seule variable

La règle de Neyman-Pearson nous donne :

Donc, à partir de , on trouve :

Après simplification de l’équation, on obtient :

Ce problème s‘exprime donc en fonction d’une seule variable, qui est x1.

### Expression de la règle en fonction de α\*

Or on a ici,

Et donc le résultat

### Frontière de décision

### Estimation de α et β

Pour estimer α, on calcule la part des vecteurs vect1X1 qui sont supérieur à la valeur de la frontière et, pour β, on calcule la part des vecteurs vect2X1 qui sont inférieur à la valeur de la frontière.

### Courbe COR 1-β =g(α\*)

Pour obtenir l’équation de la courbe, on part de l’expression suivante :

Or : donc

D’où:

Et enfin :

La représentation est donc :

### Règle de Bayes

### Expression

L’intérêt de la règle de Bayles est de prendre en compte les coûts (c) afin de minimiser le risque (π)

### Frontière de décision

### Estimation

# Conclusion