|  |
| --- |
| TP04 – SY09 |
| Analyses discriminantes quadratique et linéaire |
|  |

** Carolina BRUNO**

**Clara PAGANI**

Printemps 2013

# Régle de Bayes

Dans cette partie, nous allons utiliser la règle de Bayes afin de déterminer la frontière de décision pour 5 échantillons de données chacun suivant une distribution normale bidimensionnelle. Les paramètres variant entre les cas sont la probabilité à priori, les matrices de covariance et les centres de gravités de chacune des classes. Nous déterminerons pour chacun des ces échantillons la probabilité d’erreur observée et théorique.

Premièrement voici les calculs effectués pour chacun des cas permettant de déterminer l’équation de la frontière de décision en fonction de x1 et  :

On suppose que la population est répartie en deux classes, en proportions et , issues des distributions gaussiennes bivariées (,) et (,).

## Équation de la frontière de décision

Pour calculer les frontières de décisions, nous avons utilisé une règle d'affectation probabiliste :



Calculer le det



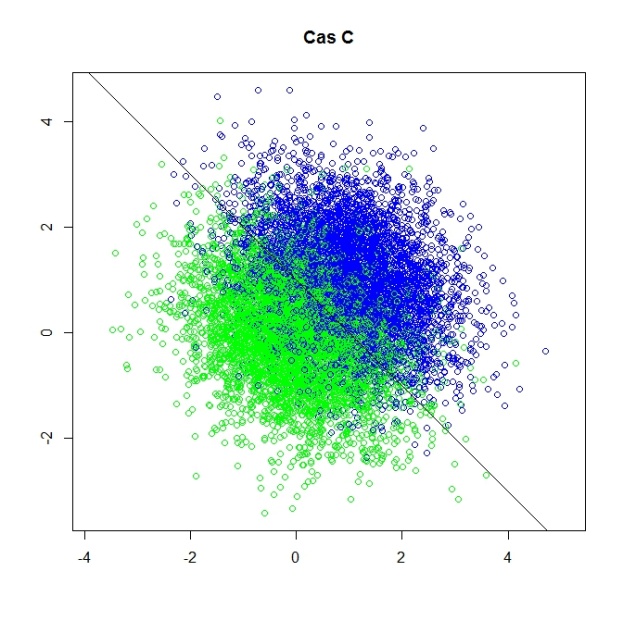
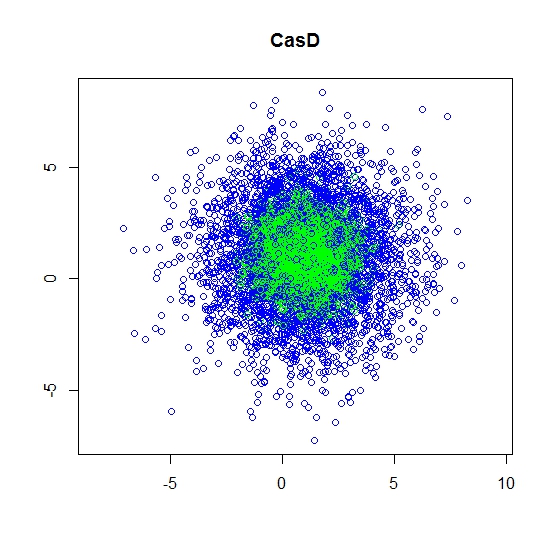
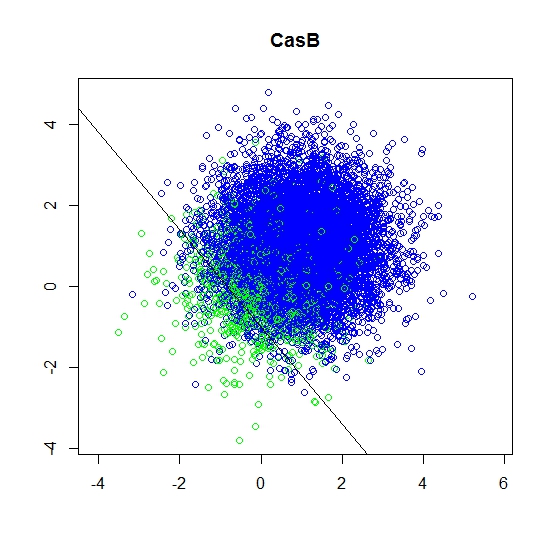
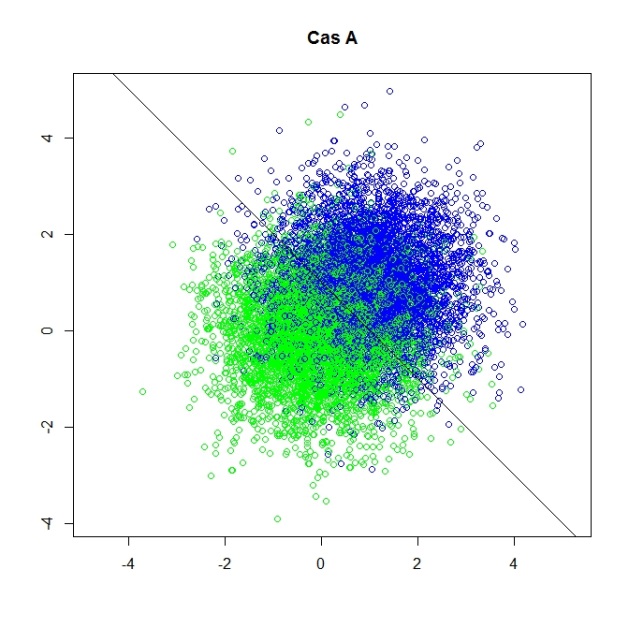
Calculer le det

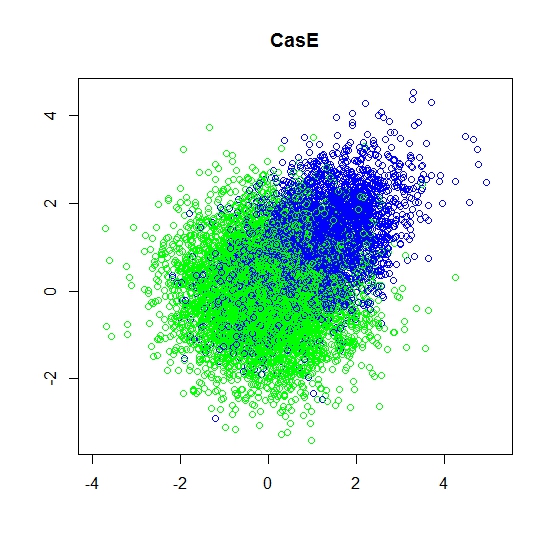


Calculer le det

## Nuage associé en représentant les deux classes de manière distincte

Nous remarquons que les 3 premiers cas ont des frontières de décisions linéaires contrairement aux deux derniers cas dont les frontières sont quadratiques. La frontière de décision du cas D est un cercle et celle du cas E est une ellipse.

Les 5 graphs ci-dessous représentent chacun des cas ainsi que leur frontière de décision pour les trois premiers cas :



**Expression de l’estimateur de probabilité d’erreur:**



**Expression de la probabilité d’erreur théorique:**

Pour 

 Distance de Mahanalobis



**Pour les autres cas, on applique la borne de Bhattacharyya :**





|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | CAS A | CAS B | CAS C | CAS D | CAS E |
| Estimation probalité d’erreur observée | 0.2410 | 0.0912 | 0.1937 | 0.2131 | 0.2313 |
| Estimation probabilité d’erreur théorique | 0.2398 | 0.0909 | 0.1990 | 0.3651 | 0.3855 |

Nous utilisons donc la distance de Mahanobis pour les cas A,B et C et la borne de Bhattacharyya pour les cas D et E.

Nous remarquons que les probabilités d’erreurs observées et théoriques des trois premiers cas sont très proches.

La différence entre ces deux probabilités est plus marquée pour les deux derniers cas. Ceci est du au fait que l’on utilise les borne de Bhattacharyya pour estimer la probabilité théorique qui maximise cette estimation.

# Analyse discriminante sur les données crabs

L’objectif de cette partie est de déterminer une frontière de décision adéquate permettant de distinguer les crabes selon leurs sexes en utilisant les paramètres FL et RW. Nous allons pour cela utiliser l’analyse discriminante linéaire et l’analyse discriminante quadratique.

## Fonctions utilisées

La **fonction *lda()*** permet d’effectuer une analyse discriminante linéaire. De même la **fonction *qda()*** permet d’effectuer une analyse discriminante quadratique.

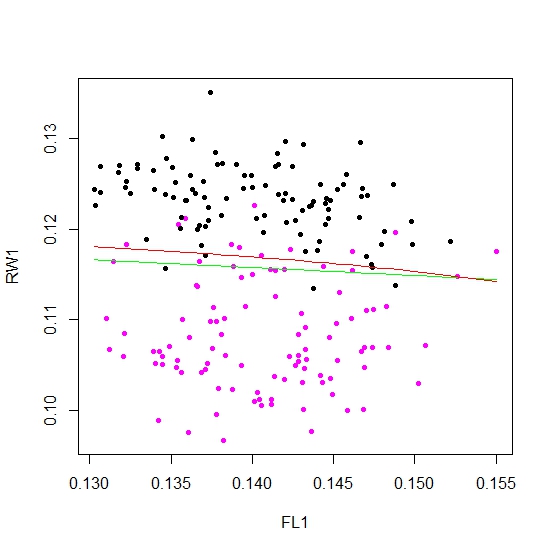
La **fonction *contour()***permet de tracer la frontière de décision, c’est-à-dire de tracer les points x tel que f(x) = 0

La **fonction *sample()*** permet de sélectionner aléatoirement n fixé éléments d’un échantillon passé en arguments.

La **fonction *predict()***permet de répartir des observations d’un échantillon dans des classes en utilisant un modèle de prédiction qui est déterminé en fonction de la classe de l’objet passé en paramètre.

La **fonction *predict.lda()***permet de répartir des observations d’un échantillon dans des classes en utilisant le modèle d’analyse discriminante linéaire.

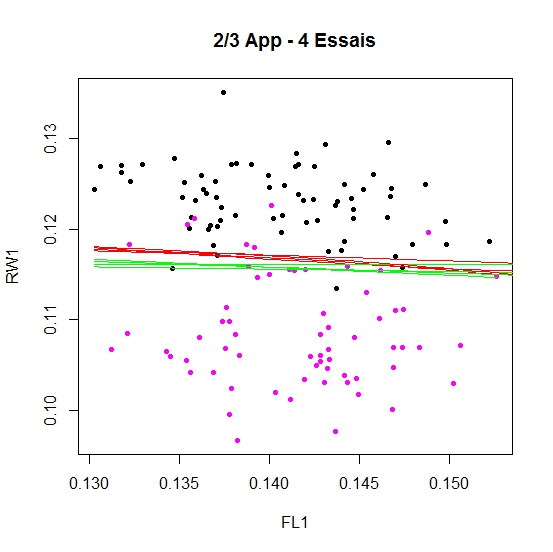
## ADL et ADQ sur l’ensemble des données crabs

Après l’application de l’analyse discriminante linéaire (ADL) et de l’analyse discriminante quadratique (ADQ) sur l’ensemble des données crabs, nous obtenons le graph ci-à gauche.

Nous distinguons en vert la frontière de décision relative à l’ADL et en rouge celle relative à l’ADQ.

Ici nous utilisons l’ensemble des données comme échantillon d’apprentissage et de test. Ci-après les résultats du calcul de la prabalité d’erreur pour chacune des méthodes :

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | ADL | ADQ |
| Erreur(%) |  | 9,5% | 8,5% |

 Nous constatons que l’ADQ présente un taux d’erreur plus faible que l’ADL. La frontière de décision déterminée par l’ADQ est donc plus précise sur ce jeu de données et permet de mieux déterminer le sexe de chacun des crabes.

## ADQ et ADL sur 2/3 des données crabs

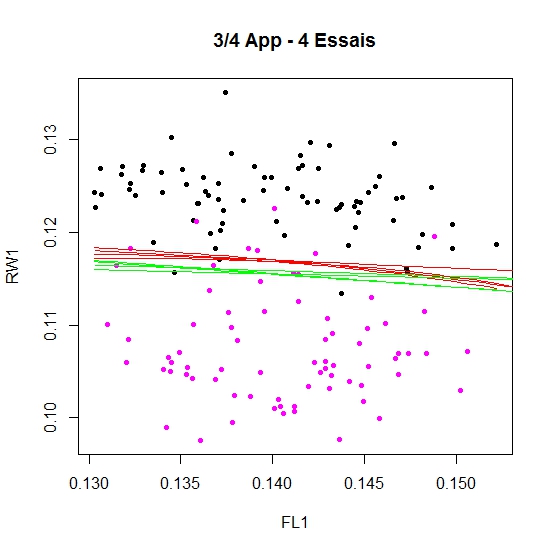
Nous effectuons la même démarche que ci-dessus excepté que notre échantillon d’apprentissage est composé de 2/3 des données crabs sélectionnées aléatoirement parmi l’ensemble des données crabs.

Notre échantillon test sera le tiers restant.

Nous répétons cette démarche 4 fois.

Nous calculons ensuite pour chacun des échantillons (apprentissage et test) la moyenne de probabilité d’erreur.

Nous ré itérons ensuite tous ce processus mais pour un échantillon d’apprentissage est composé de 3/4 des données.

Nous obtenons les deux graphes ci à droite et le tableau de résultat ci-dessous :

|  |  |  |
| --- | --- | --- |
|  | ADL | ADQ |
| Moy Prob erreur estimée ech. App (2/3) | 8,45% | 7,71% |
| Moy Prob erreur estimée ech. test.(1/3) | 11,6% | 9,70% |
|  | | |
| ADL | **ADQ** |
| Moy Prob erreur estimée ech. App (3/4) | 9,67% | 8,33% |
| Moy Prob erreur estimée ech. test.(1/4) | 8.5% | 7,00% |

Premièrement nous remarquons que plus

Nous remarquons également que dans tous les cas la probabilité de l’erreur de l’analyse discriminante quadratique est inférieure à celle de l’Analyse discriminante linéaire

# Exercice

## Paramètres du modèle sous chacune des hypothèses

## Classifieurs à la règle de Bayes

## Règle de décision