

# Rapport du TP3 de MA 412

Paul Ezvan et Omar Givernaud

9 novembre 2009

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Représentation des données</b>	<b>3</b>
2.1	Normalisation . . . . .	3
2.2	Projection . . . . .	3
2.3	Représentation . . . . .	4
<b>3</b>	<b>Discrimination par la règle des k-ppv</b>	<b>5</b>
<b>4</b>	<b>Discrimination Bayésienne</b>	<b>6</b>
4.1	Données normalisées . . . . .	6
4.2	Données normalisées projetées . . . . .	7
<b>5</b>	<b>Conclusion</b>	<b>10</b>

# 1 Introduction

Ce TP a pour but de mettre en œuvre les méthodes de classification vues en cours.

## 2 Représentation des données

### 2.1 Normalisation

Nous dûmes modifier quelque peu notre fonction de normalisation afin de pouvoir récupérer les moyennes et écarts types utilisés pour normaliser X et normaliser Xc avec ceux-ci.

```
1 function [ Y, M, E ] = normalize(X, Ms, Es)
2 %UNTITLED3 Summary of this function goes here
3 % Detailed explanation goes here
4 if (nargin < 3)
5     M=zeros(size(X,2),1);
6     E=zeros(size(X,2),1);
7     for i=1:size(X,2)
8         Y(:,i)=(X(:,i)-mean(X(:,i)))/std(X(:,i),1);
9         M(i)=mean(X(:,i));
10        E(i)=std(X(:,i),1);
11    end
12 else
13     for i=1:size(X,2)
14         Y(:,i)=(X(:,i)-Ms(i))/Es(i);
15     end
16 end
```

### 2.2 Projection

Nous représentâmes X à l'aide de la fonction classr puis nous projetâmes Xc normalisé sur le plan principal.

```
1 load Emploi_eur_disc
2 % Normalisation des données
3 [Y,M,E]=normalize(X)
4 Yc=normalize(Xc,M,E)
5 % Calcul des axes principaux
6 R=corrcoef(Y);
7 [V, D]=eig(R);
8 [Ld, I]=sort(diag(D));
9 Vd = V(:,I)
10 % Projection dans le plan principal de Xc
```

```

11 A=[Vd(:,size(Yc,2)),Vd(:,size(Yc,2)-1)]
12 P=Yc*A
13 figure
14 hold on
15 % Représentation des données déjà classées
16 classr(Y,L,q)
17 % Représentation de Xc
18 plot(P(:,1),P(:,2),'x')
19 text(P(:,1),P(:,2),qc)

```

## 2.3 Représentation

Voici le résultat obtenu.

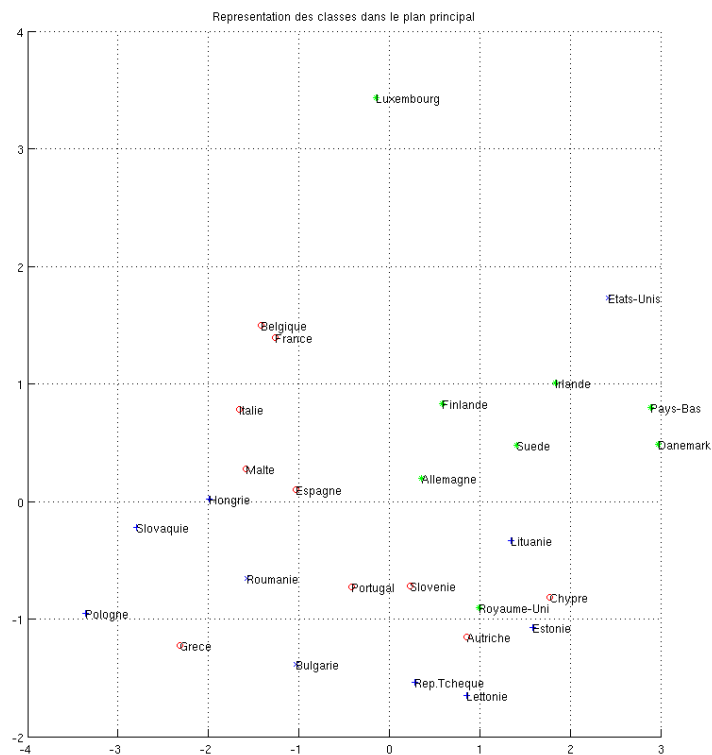


FIGURE 1 – Représentation des données classifiées et des données à discriminer

### 3 Discrimination par la règle des k-ppv

Les 3 options 'nearest', 'random' et 'consensus' correspondant à la règle de discrimination.

- nearest : Majorité, choix de la classe du plus proche voisin en cas d'égalité.
- random : Majorité, choix tiré au sort de manière uniforme en cas d'égalité.
- consensus : consensus requis sinon pas de classement.

```
1 >> Lc=knnclassify (Yc,Y,L,1,'euclidean','consensus')
2 Lc =
3     2     1     3
4 >> Lc=knnclassify (Yc,Y,L,2,'euclidean','consensus')
5 Lc =
6     NaN    NaN     3
7 >> Lc=knnclassify (Yc,Y,L,3,'euclidean','consensus')
8 Lc =
9     NaN    NaN     3
10 >> Lc=knnclassify (Yc,Y,L,4,'euclidean','consensus')
11 Lc =
12     NaN    NaN     3
13 >> Lc=knnclassify (Yc,Y,L,1,'euclidean','random')
14 Lc =
15     2     1     3
16 >> Lc=knnclassify (Yc,Y,L,2,'euclidean','random')
17 Lc =
18     1     1     3
19 >> Lc=knnclassify (Yc,Y,L,3,'euclidean','random')
20 Lc =
21     1     2     3
22 >> Lc=knnclassify (Yc,Y,L,4,'euclidean','random')
23 Lc =
24     1     2     3
25 >> Lc=knnclassify (Yc,Y,L,1,'euclidean','nearest')
26 Lc =
27     2     1     3
28 >> Lc=knnclassify (Yc,Y,L,2,'euclidean','nearest')
29 Lc =
30     2     1     3
31 >> Lc=knnclassify (Yc,Y,L,3,'euclidean','nearest')
32 Lc =
33     1     2     3
34 >> Lc=knnclassify (Yc,Y,L,4,'euclidean','nearest')
35 Lc =
36     1     2     3
```

Nous remarquons que la discrimination du troisième point semble aisée puisqu'il est discriminé avec la méthode consensus même avec quatre ppv. Les méthodes random et nearest donnent le même résultat avec un nombre élevé de ppv, mais différent avec deux ppv.

## 4 Discrimination Bayésienne

Classify permet de réaliser une discrimination par mesure de voisinage de Mahalanobis ou discrimination bayésienne. Linear effectue une séparation linéaire, quadratic une séparation quadratique et mahalanobis une séparation à l'aide de la distance de Mahalanobis.

### 4.1 Données normalisées

```
1 >> [CLASS,ERR,POSTERIOR]=classify(Yc,Y,L,'linear')
2 CLASS =
3     1     1     3
4 ERR =
5     0.1000
6 POSTERIOR =
7     0.9831     0.0169     0.0000
8     0.9832     0.0168     0.0000
9     0.0000     0.0004     0.9996
10
11 >> [CLASS,ERR,POSTERIOR]=classify(Yc,Y,L,'quadratic')
12 CLASS =
13     2     2     3
14 ERR =
15     0
16 POSTERIOR =
17     0.0000     1.0000     0.0000
18     0.0000     1.0000     0.0000
19     0.0000     0.0001     0.9999
20
21 >> [CLASS,ERR,POSTERIOR]=classify(Yc,Y,L,'mahalanobis')
22 CLASS =
23     2     2     3
24 ERR =
25     0.0476
26 POSTERIOR =
27     []
```

Les meilleurs taux d'erreur sont obtenus avec la méthode quadratique, qui abouti à la même classification que celle obtenue en utilisant la distance de Mahalanobis.

## 4.2 Données normalisées projetées

Nous réitérâmes cette discrimination sur les données projetées.

```
1 >> [CLASS,ERR,POSTERIOR]= classify (Pc ,P,L, ' linear ')
2 CLASS =
3     1     1     3
4 ERR =
5     0.3452
6 POSTERIOR =
7     0.6992     0.2973     0.0034
8     0.5787     0.4140     0.0073
9     0.0008     0.0084     0.9908
10
11 >> [CLASS,ERR,POSTERIOR]= classify (Pc ,P,L, ' quadratic ')
12 CLASS =
13     1     1     3
14 ERR =
15     0.3452
16 POSTERIOR =
17     0.6650     0.3307     0.0043
18     0.6458     0.3516     0.0026
19     0.0000     0.0008     0.9992
20
21 >> [CLASS,ERR,POSTERIOR]= classify (Pc ,P,L, ' mahalanobis ')
22 CLASS =
23     1     1     3
24 ERR =
25     0.3452
26 POSTERIOR =
27     []
```

Cette fois-ci les discrimination obtenues et les taux d'erreur sont identiques avec les trois méthodes. Le taux d'erreur est plus important que celui obtenu avec les données originelles et la discrimination obtenue est différente.

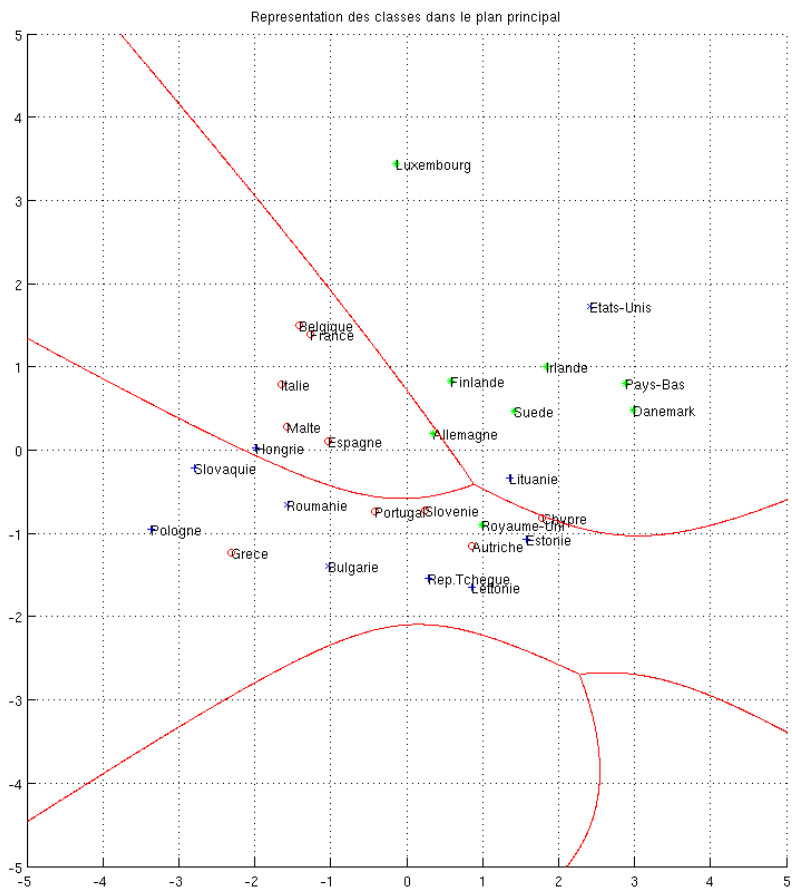


FIGURE 2 – Courbe séparatrice avec la distance de Mahalanobis



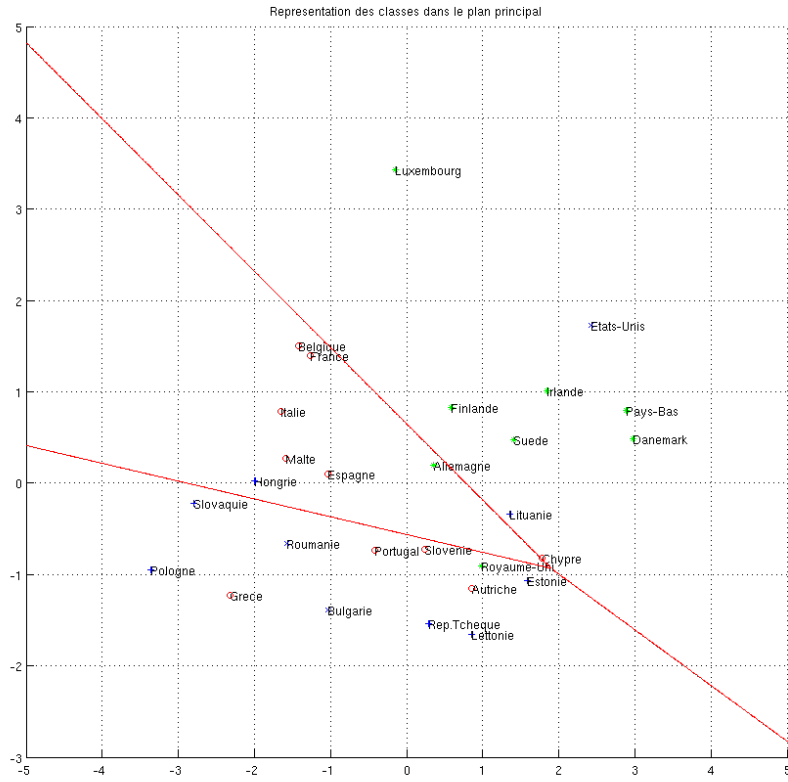


FIGURE 3 – Courbe séparatrice linéaire

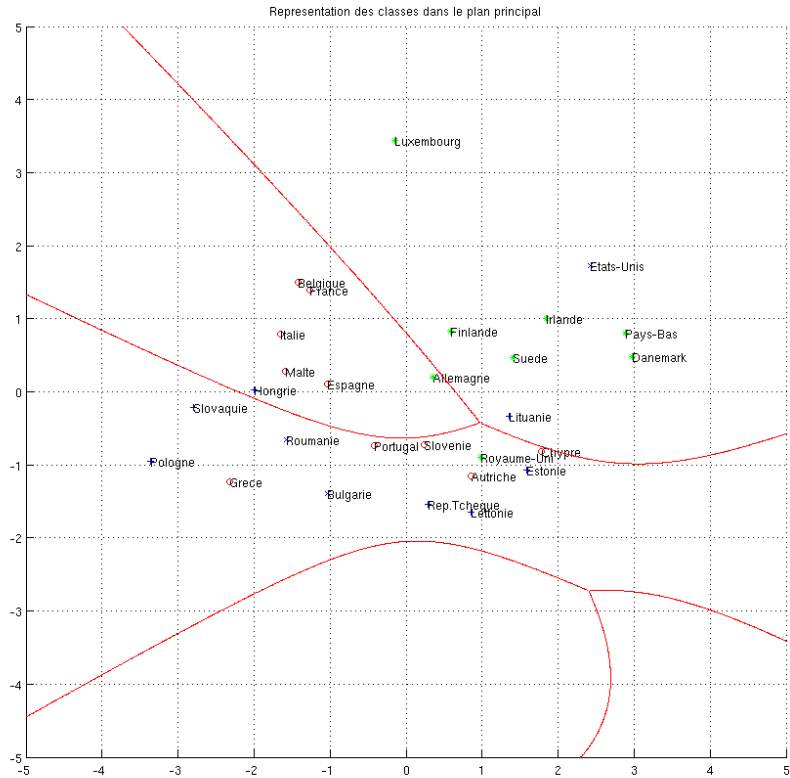


FIGURE 4 – Courbe séparatrice quadratique

Les courbes obtenues avec la méthode quadratique et avec la distance de Mahalanobis sont similaires. La discrimination est la même avec les trois méthodes.

## 5 Conclusion

Ce TP nous permet de comparer les différentes méthodes de discrimination. Nous observâmes que la discrimination Bayésienne à l'aide de la méthode quadratique est très proche de la discrimination par distance de Mahalanobis.