

1 Exercices

1. Calculer la règle de discrimination du maximum de vraisemblance sur la base d'observations unidimensionnelle de distribution exponentielle.

La fonction de densité de la distribution exponentielle $Exp(\lambda)$ est:

$$f(x) = \lambda \exp\{-\lambda x\} \text{ pour } x > 0.$$

On compare les vraisemblances des deux populations $\Pi_1 = Exp(\lambda_1)$ et $\Pi_2 = Exp(\lambda_2)$, on affecte l'observation x à la population Π_1 si

$$\begin{aligned} L_1(x) &\geq L_2(x) \\ L_1(x)/L_2(x) &\geq 1 \\ \frac{\lambda_1}{\lambda_2} \exp\{-x(\lambda_1 - \lambda_2)\} &\geq 1 \\ \log \frac{\lambda_1}{\lambda_2} - x(\lambda_1 - \lambda_2) &\geq 0 \\ x(\lambda_1 - \lambda_2) &\leq \log \frac{\lambda_1}{\lambda_2}. \end{aligned}$$

En supposant que $\lambda_1 < \lambda_2$, on obtient la règle discriminante:

$$R_1 = \left\{ x : x \geq \frac{\log \lambda_1 - \log \lambda_2}{\lambda_1 - \lambda_2} \right\}$$

L'observation x est classifiée dans la population Π_1 si sa valeur est plus grande que la constante $(\log \lambda_1 - \log \lambda_2)/(\lambda_1 - \lambda_2)$.

2. Montrer que le discriminant linéaire de Fisher est équivalent à la règle du maximum de vraisemblance pour des distributions multivariées normales avec des matrices de covariances égales ($J=2$).

On a vu en cours que la règle du MV s'écrit

$$R_1^{ML} = \{x : \alpha^T(x - \mu) \geq 0\},$$

où $\alpha = \Sigma^{-1}(\mu_1 - \mu_2)$ et $\mu = \frac{1}{2}(\mu_1 + \mu_2)$, et que la règle du discriminant linéaire de Fisher pour $J = 2$ est:

$$R_1^F = \{x : (\bar{x}_1 - \bar{x}_2)^T \mathcal{W}^{-1}(x - \bar{x}) \geq 0\},$$

on a également vu que $\mathcal{W} = n\mathcal{S}$, où \mathcal{S} est la matrice de covariance et n le numéro d'observations. En définissant la version empirique de α comme $\hat{\alpha} = (\bar{x}_1 - \bar{x}_2)\mathcal{S}^{-1}$, on peut réécrire la règle du discriminant de Fisher comme:

$$R_1^F = \{x : \hat{\alpha}^T(x - \bar{x}) \geq 0\}.$$

En comparant cette expression avec la règle discriminante du MV, on voit que la règle de Fisher R_1^F peut être interprétée comme la version empirique de l'estimation de la règle discriminante du MV R_1^{ML} .

3. On suppose que les observations proviennent de trois populations distinctes, Π_1, Π_2 et Π_3 , caractérisées par leur distributions binomiales:

$\Pi_1 : X \sim Bi(10, 0.2)$ avec la probabilité à priori $\pi_1 = 0.5$;

$\Pi_2 : X \sim Bi(10, 0.3)$ avec la probabilité à priori $\pi_2 = 0.3$;

$\Pi_3 : X \sim Bi(10, 0.5)$ avec la probabilité à priori $\pi_3 = 0.2$.

Utiliser la méthode de Bayes pour déterminer les règles discriminantes R_1, R_2 et R_3

Les règles discriminantes de Bayes correspondantes R_j pour $j = 1, 2, 3$ sont définies par:

$$R_j = \{x \in \{0, 1, \dots, 9, 10\} : \pi_j f_j(x) \geq \pi_i f_i(x) \text{ pour } i = 1, 2, 3\}.$$

Les valeurs de $\pi_i f_i(x)$, pour $i = 1, \dots, 3$ et $x = 0, \dots, 10$ sont données dans la table 1 dont on déduit directement les règles discriminantes:

$$R_1 = \{0, 1, 2, 3\},$$

$$R_2 = \{4\},$$

$$R_3 = \{5, 6, 7, 8, 9, 10\}.$$

x	$f_1(x)$	$f_2(x)$	$f_3(x)$	$\pi_1 f_1(x)$	$\pi_2 f_2(x)$	$\pi_3 f_3(x)$	$\pi_j f_j(x)$	j
0	0.107374	0.028248	0.000977	0.053687	0.008474	0.000195	0.053687	1
1	0.268435	0.121061	0.009766	0.134218	0.036318	0.001953	0.134218	1
2	0.301990	0.233474	0.043945	0.150995	0.070042	0.008789	0.150995	1
3	0.201327	0.266828	0.117188	0.100663	0.080048	0.023438	0.100663	1
4	0.088080	0.200121	0.205078	0.044040	0.060036	0.041016	0.060036	2
5	0.026424	0.102919	0.246094	0.013212	0.030876	0.049219	0.049219	3
6	0.005505	0.036757	0.205078	0.002753	0.011027	0.041016	0.041016	3
7	0.000786	0.009002	0.117188	0.000393	0.002701	0.023438	0.023438	3
8	0.000074	0.001447	0.043945	0.000037	0.000434	0.008789	0.008789	3
9	0.000004	0.000138	0.009766	0.000002	0.000041	0.001953	0.001953	3
10	0.000000	0.000006	0.000977	0.000000	0.000002	0.000195	0.000195	3

Figure 1: Les valeurs de la vraisemblance et de la vraisemblance de Bayes pour les trois distributions binomiales.

4. Soit le jeux de données suivant :

	t=1			t=2		
x_1	2	4	3	5	3	4
x_2	12	10	8	7	9	5

x_1 et x_2 suivent une distribution normale bivariée au sein de chaque groupe, où la matrice de covariance matrix est supposée être la même dans les deux groupes.

- Estimer la moyenne des groupes, la matrice de covariance, et la probabilité à priori à partir de cette ensemble de données d'apprentissage.

La moyenne des groupes :

$$\hat{\mu}_j = \frac{1}{n_j} \sum_{y_i=j} x_j, \text{ moyenne classe } j$$

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 3 \\ 10 \end{bmatrix} \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$$

la matrice de covariance commune:

$$\hat{\Sigma} = \frac{1}{n-K} \sum_{j=1}^K (x_i - \hat{\mu}_j)(x_i - \hat{\mu}_j)^T$$

(où n_j est le nombre de point de la classe j)

$$\hat{\Sigma} = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$$

la probabilité à priori $\hat{\pi}_j = n_j/n$ proportion d'observation pour la classe j

$$\hat{\pi}_{j=1,2} : p(C_1) = p(C_2) = 3/6 = 1/2$$

- Estimer les fonctions discriminantes linéaires $a_1(x_1, x_2)$ and $a_2(x_1, x_2)$ pour les groupes 1 et 2 respectivement.

Fonctions de discrimination estimées :

$$\hat{\delta}_j(x) = x^T \hat{\Sigma}^{-1} \hat{\mu}_j - \frac{1}{2} \hat{\mu}_j^T \hat{\Sigma}^{-1} \hat{\mu}_j + \log \hat{\pi}_j$$

$$\begin{aligned} a_1(\mathbf{x}) &= \bar{\mathbf{x}}_1^T \hat{\Sigma}^{-1} \mathbf{x} - \frac{1}{2} \bar{\mathbf{x}}_1^T \hat{\Sigma}^{-1} \bar{\mathbf{x}}_1 + \ln \frac{1}{2} \\ &= \frac{1}{3} \begin{bmatrix} 3 & 10 \end{bmatrix} \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \frac{1}{6} \begin{bmatrix} 3 & 10 \end{bmatrix} \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 10 \end{bmatrix} + \ln \frac{1}{2} \\ &= 7\frac{1}{3}x_1 + 4\frac{1}{3}x_2 - 32\frac{2}{3} + \ln \frac{1}{2} \end{aligned}$$

$$\begin{aligned} a_2(\mathbf{x}) &= \bar{\mathbf{x}}_2^T \hat{\Sigma}^{-1} \mathbf{x} - \frac{1}{2} \bar{\mathbf{x}}_2^T \hat{\Sigma}^{-1} \bar{\mathbf{x}}_2 + \ln \frac{1}{2} \\ &= \frac{1}{3} \begin{bmatrix} 4 & 7 \end{bmatrix} \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \frac{1}{6} \begin{bmatrix} 4 & 7 \end{bmatrix} \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 7 \end{bmatrix} + \ln \frac{1}{2} \\ &= 7\frac{2}{3}x_1 + 3\frac{2}{3}x_2 - 28\frac{1}{6} + \ln \frac{1}{2} \end{aligned}$$

- Donner une règle de classification linéaire pour ce problème et construisez une matrice de confusion en appliquant cette règle à l'ensemble d'apprentissage. Quelle est le taux d'erreur ?

Les fonctions discriminantes linéaires

$$\begin{aligned} \hat{\delta}_j(x) &= x^T \hat{\Sigma}^{-1} \hat{\mu}_j - \frac{1}{2} \hat{\mu}_j^T \hat{\Sigma}^{-1} \hat{\mu}_j + \log \hat{\pi}_j \\ &= a_j + b_j^T x \end{aligned}$$

sont des fonctions affines de x .

La frontière de décision entre les classes j, k est l'ensemble de tous les $x \in \mathbf{R}^p$ tel que $\hat{\delta}_j(x) = \hat{\delta}_k(x)$, c'est à dire

$$a_j + b_j^T x = a_k + b_k^T x$$

ce qui définit le sous-espace affine en x suivant :

$$a_j - a_k + (b_j - b_k)^T x = 0$$

Ce qui donne sur l'exemple :

$$a(\mathbf{x}) = a_1(\mathbf{x}) - a_2(\mathbf{x}) = -\frac{1}{3}x_1 + \frac{2}{3}x_2 - 4\frac{1}{2}$$

Si $a(\mathbf{x}) < 0$ on affecte au groupe 2, autrement au groupe 1.

Matrice de confusion (ligne : affectation, colonne : réalité)

	groupe 1	groupe 2
groupe 1	2	1
groupe 2	1	2

Figure 2: Matrice de confusion.

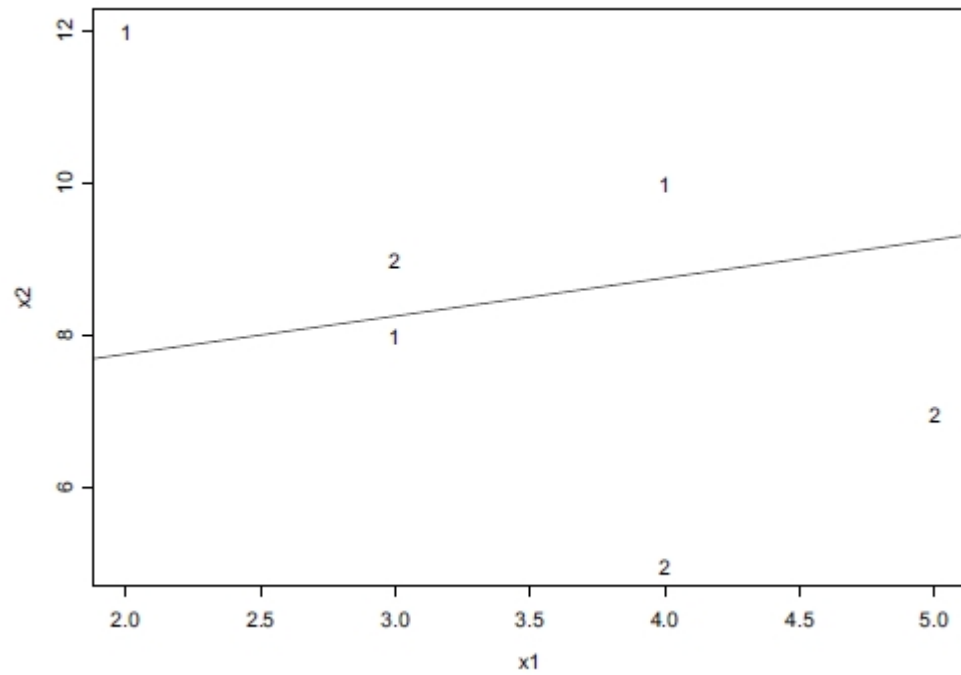
Donc l'erreur apparente est de $1/3$

- Dessiner la frontière entre les zones qui (selon la fonction de classification calculée précédemment) appartiennent au groupe 1 et 2 respectivement, dans un "scatterplot" des données (diagramme de dispersion). Pouvez-vous trouver une ligne droite qui a le plus faible taux d'erreur?

La frontière entre les deux zones est donnée par $x_2 = 6\frac{3}{4} + \frac{1}{2}x_1$:

On peut tracer une ligne qui conduise à ne commettre qu'une seule erreur

Figure 3: Frontière de décision



```
myData <- data.frame(x1=c(2,4,3,5,3,4),
x2=c(12,10,8,7,9,5),group=c(1,1,1,2,2,2))

plot(myData[myData$group==1,-3]$x1,
myData[myData$group==1,-3]$x2,xlim=c(1,5),ylim=c(1,12),
xlab="x1",
ylab="x2")
points(myData[myData$group==2,-3]$x1,myData[myData$group==2,-3]$x2,
pch=19)
```

```
g1 <- myData[myData$group==1,-3]
g2 <- myData[myData$group==2,-3]
```

```
#Calcul des vecteurs moyennes, des matrices de covariance
```

```

#et de la matrice de covariance commune
(pooled covariance matrix):

m1 <- as.vector(apply(g1, 2, mean))
m2 <- as.vector(apply(g2, 2, mean))

s1 <- cov(g1)
s2 <- cov(g2)

s <- (s1+s2)/2
#Inverse matrice de covariance commune
si <- solve(s)

a <- t(m1-m2)%*%si
m <- t(m1-m2)%*%si%*(m1+m2)/2 + log(0.5/0.5)

# On trace la frontière de décision
plot(myData[myData$group==1,-3]$x1,myData[myData$group==1,-3]$x2,
xlim=c(1,5),ylim=c(1,12),
      xlab="x1",
      ylab="x2")
points(myData[myData$group==2,-3]$x1,myData[myData$group==2,-3]$x2,
pch=19)
text(110,22,"R1",col="blue")
text(30,15,"R2",col="blue")
abline(m/a[2],-a[1]/a[2],lwd=2,col="blue")

```