Chapitre 2 Méthodes probabilité

Youming

2023-09-13

背景 Constexte et objectifs

有 n 组 $(x_i, z_i)\backslash$ x_i 是变量\ z_i 是组，若 $z_i = 1$，则 x_i 在 G 组。\

目的 L'objectif

Affecter à l'une des classes un nouvel individu $X_{n+1} \in X$ dont on ignore z_{n+1}

方法 La méthode

Définir `G` scores par chaque $S_g : x \to R$ permettant d'affecter x_{n+1} selon une règle du type (par maximiser de score)

$$\hat{z}_{n+1,g} = 1 \Leftrightarrow S_g(X_{n+1}) = \max\{S_j(X_{n+1}); j = 1, ..., G\}$$

Analyse Discriminante (AD) probabiliste

Modèle : On supose que :

- $\{x, z\}$ 互相独立:les coupls (x_i, z_i) sont des réalisation indépendants de $(x, z) \in X_{x\{0;1\}^G}$
- $z \sim M_G(1; \pi_1, \pi_2 ... \pi_G)$
  - 他的参数 (essai de paramètre) 是：$\{\pi_1, \pi_2 ... \pi_G\}$

    $\pi_g > 0$ 且 $\sum_{g=1}^{G} z_g = 1$
  - 以 N = 1 和 $\{\pi_1, \pi_2 ... \pi_G\}$ 为参数的 La loi Multi-Normale à l'ordre de G $M_G(N; \pi_1, \pi_2 ... \pi_G)$

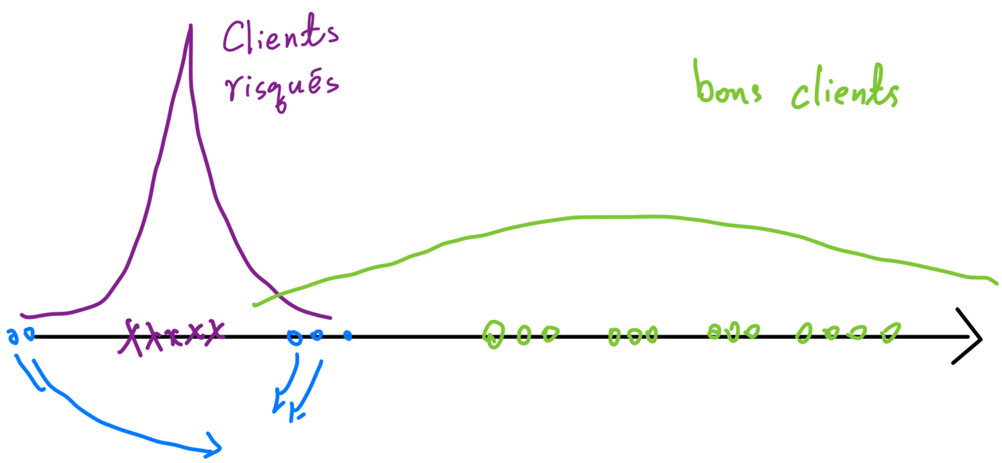
- 在给定条件 $z_i g = 1$ 下，条件向量 x 的分布是由一个以 x 为支持（即可能取值的范围）的概率分布所描述，这个概率分布的概率密度函数是 $f_g(*;\theta_g)$，其中 θ_g 是分布的参数。
  - 概率密度函数 $f_g(*;\theta_g)$ 通常属于相同的概率分布家族，选择概率分布家族的具体形式可能会依赖于数据的性质。
- $y_i = \alpha n_i + b + \varepsilon$ d'où $(\varepsilon \sim N(0,1))$
- $z \sim M_G(1;\pi_1,\pi_2...\pi_G)$ −>Loi multinomiale d'ordre G de paramètre $N = 1et\ \pi = (\pi_1;....;\pi_G) \Rightarrow M(N;(\pi_1;....;\pi_G)$

  le 1 c'est une individu qu'on choisir et c'est une compossante, est tous les composant vaut 0
- $(x|Z_y = 1) \sim f_g(*;\theta_g) =$ Fonction de probabilité de descropteur dans la classe g

  Pappel : $y = (y_1;...;y_k)' \sim M_k(N;\pi_1,...,\pi_K)$ sinifie :
  - y prend se valeur dans $\{(n_1;n_2,...,n_k)' \in N^K$

    $\sum_{k=1}^{K} = N$

$$P(y = (n_1, n_2, ..., n_k)') = \frac{N!}{n_1!n_2!...n_k!\Pi_{j=1}^{k}\pi_j n_j}$$



Dans la période risquée, il y a des clients atypique

Donc on préfère **loi de student** parce que elle est plus épais que ceci.

La méthode :

- On va estimer la paramétrie θ

  Maximum de vraisemblance 最大似然估计

  $\theta = (\pi_1, ...\pi_\theta, \theta_1, ..., \theta_G)$

  - $\pi_1, ...\pi_\theta$: poids de classe
  - π_g : proba d'appartien à la classe g

    $\pi_g = P(z_g = 1)$
  - $\theta_1, ..., \theta_G$: Parametre des fonction probabilité conditionnelle 条件概率函数的参数.

理解:

- 不分组
  - 分布依赖同一个参数 θ

    似然数:

    $$L(\theta) = \Pi_{j=1}^{n} P(x_j \in A_j|\theta)$$

    对数似然数

    $$l(\theta) = \sum_{j=1}^{n} \ln f_x(x_j|\theta)$$
- 分组数据情况，也就是这节课的情况:
  - 选出一组数据，现成题中的数据。$\{c_0, c_1, ..., c_j\}$
  - n_j 是区间 $(c_j - 1, c_j]$ 的观测数目。

    似然数:

    $$L(\theta) = \Pi_{j=1}^{k} \Big[F(c_j|\theta) - F(c_{j-1}|\theta]) \Big)]^{z_{i,g}}$$
  - 老师的版本，其实是一样的: $L(\theta) = \Pi_{i=1}^{n} \Pi_{i=1}^{G} \Big[\pi_g f_g(i; \theta_g) \Big]^{z_{i,g}}$
  - $l(\theta) = \ln(L(\theta))$ ，老师的是 $l(\theta) = \log(L(\theta))$ 细细品
- 令 $\frac{\partial l(\theta)}{\partial \theta} = 0$ 得到 θ.

Par exemple:

Si la classe g est gaussienne,

$$\theta_y = (\underbrace{\mu_g}_{centre\ de\ classe\ g,\ covarrience\ de\ class}; \sum g)$$

Dans ce cours , on estime par Maximiser de Vraisemblence (MV)

$$\hat{\theta} = \arg\max_{\theta} \left\{ l\Big(\theta; \{x_i, x_z\}\Big) = log\Big\{\Pi_{i=1}^{n} \Pi_{g=1}^{G} [\pi_g f_g(x_i; \theta_g)]^{z_{i,g}} \Big\} \right\}$$

On estime la Probabilité pour x_{n+1} de partenir à la classe g par le score :

$$\Pi(Z_{n+1,g} = 1 | X_{n+1}) = t_g(X_{n+1}) = \frac{\hat{\Pi}_g f_g(X_{n+1}, \hat{\theta}_g)}{\sum_{j=1}^{G} \hat{\Pi}_j f_j(X_{n+1}; \hat{\theta}_j)}$$

Explication

- Que représente le score, la variable $z_{n+1,g} \in \{0, N\}$ suite **une lois de Bernoilli** de parametre :

- Rapelle :

$$P(X = x) = \begin{cases} p & \text{si } x = 1 \\ 1 - p & \text{si } x = 0 \end{cases}$$

$$P(X - x) = p^x (1-p)^{1-x}; x \in \{0, 1\}$$

$$\Pi(z_{n+1,g} = 1 | X_{n+1}) = \frac{\Pi(X_{n+1} | Z_{n+1,g} = 1) \times \Pi(Z_{n+1,g} = 1)}{\Pi(x_{n+1})}$$

$$= \frac{\Pi(X_{n+1} | Z_{n+1,g} = 1) \times \Pi(Z_{n+1,g} = 1)}{\sum_{j=1}^{G} \Pi(X_{n+1} | Z_{n+1,g} = 1) \Pi(Z_{n+1,g} = 1)}$$

$$= \frac{f_g((X_{n+1} \hat{\theta}_g) \hat{\theta}_g}{\sum_{j=1}^{G} f_i((X_{n+1} \hat{\theta}_j)}$$

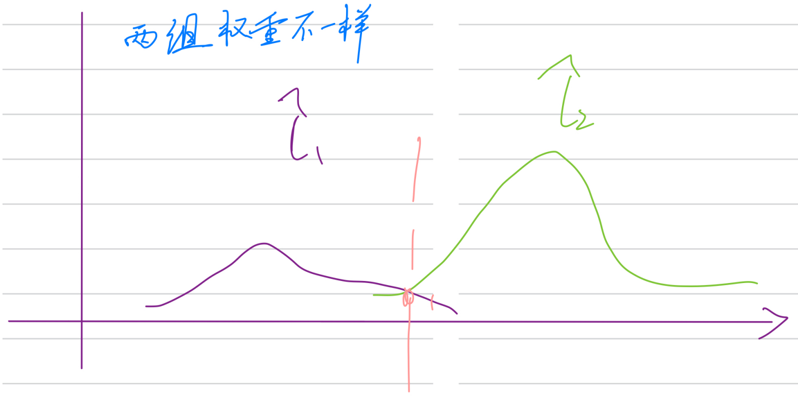
$$= E(Z_{n+1,g} | X_{n+1})$$

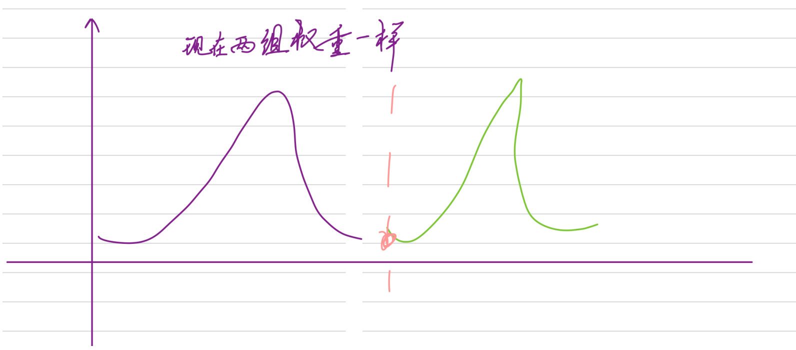
- On affecte X_{n+1} à la classe g dans laquelle il a le plus de chance de se trouver :

$$\hat{Z}_{n+1} = g \Leftrightarrow \forall_j \in \{1, ..., G\} \qquad Z_j(X_{n+1}) \leq Z_g(X_{n+1})$$

Choix de modèle :

La classe estimer de x_{n+1} dépend des fonctions de probabilité conditionnelle $f_g(\bullet;\theta_g)$ choisi et de l'espace Θ du parametre θ





Comment choisir un modèle ?

- Le créteurs d'information 赤池信息准则（AIC）, 贝叶斯信息准则（BIC）,... (propres aux modèle paramétrique au semi-problématique)

Dans ce cours

$$BIC : -2l\left(\hat{\theta};\{(x_i;z_i\}\right) + \frac{\eta}{2}\log(n)$$
$$AIC : -2l\left(\hat{\theta};\{(x_i;z_i\}\right) + \eta$$

où $\eta =$ dimension du paramètre 模型的自由参数个数 (nombre de corspante libre alqélquiloriquement)

NB : La vrai définition de BIC $= -2l\left(\hat{\theta};\{(x_i;z_i\}\right) + \eta\log(n)$

Il y a des version différent dans vertain cours.

- BIC

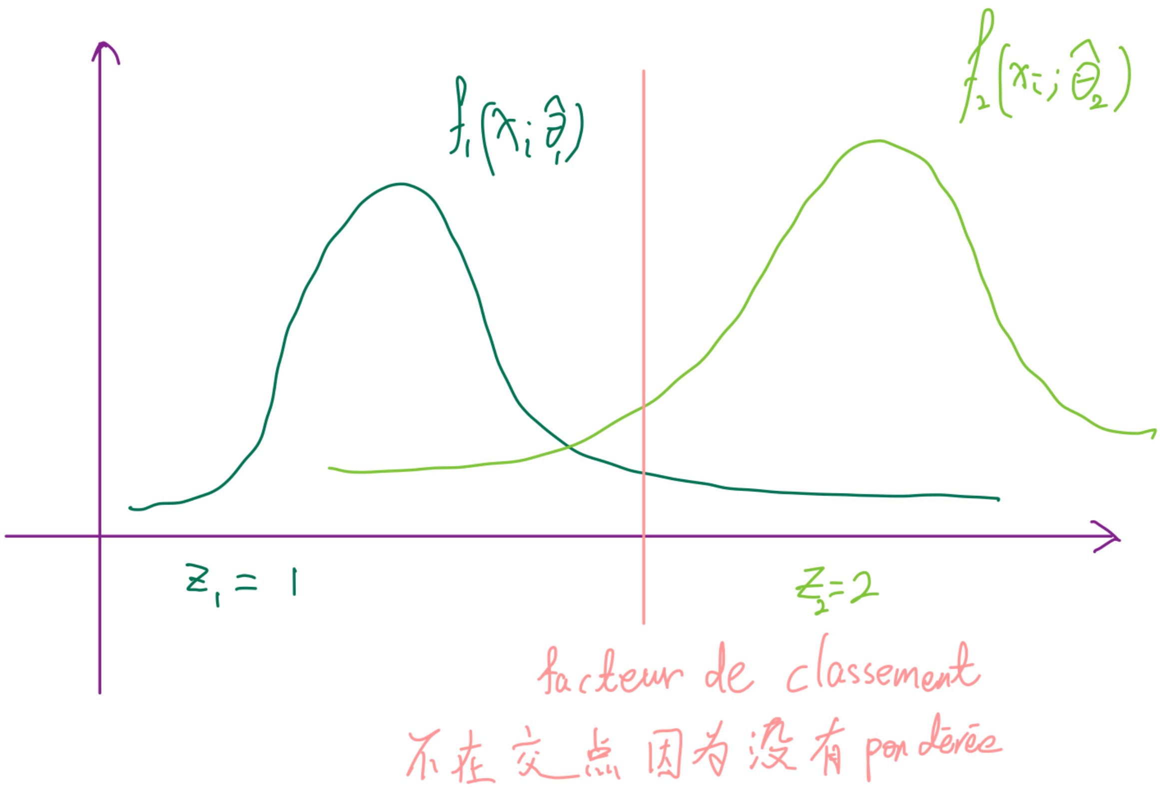
倾向于选择拟合较好但参数较少的模型，BIC 的应用通常是比较不同模型的 BIC 值，选择具有最低 BIC 值的模型作为最优模型。BIC 在避免过拟合的同时鼓励选择具有良好拟合能力的模型。注意：在小样本情况下，BIC 更倾向于选择参数更少的模型，而在大样本情况下，它更倾向于选择能够更好拟合数据的模型。因此，在使用 BIC 时，需要根据具体问题和数据规模谨慎选择。

- AIC

  与 BIC 不同，AIC 的惩罚项只包含了 2 倍自由参数的数量，而不像 BIC 那样与自由参数的数量成正比。这使得 AIC 在模型选择时更倾向于选择包含稍多一些参数但能更好拟合数据的模型。选择具有最低 AIC 值的模型作为最优模型。

L'erreur de classement

$$\hat{\varepsilon} = \sum_{g=1}^{G} \hat{\Pi}_g \int_{\{x \in X; \hat{Z}_g \neq 1\}} f_g(X_i; \hat{\theta}_g) dx$$



Justification de la formule avec G =2 classe

$$
\begin{aligned}
\hat{\varepsilon} &= \pi(erreur) \\
&= \sum_{g=1}^{G} (Classer\ x\ dans\ g\ et\ x\ n'appartient\ pas\ \ g) \\
&= \pi\Bigg(\Big[(\hat{z}_2 = 1) \cap (z_1 = 1) \Big] \cup \Big[(\hat{z}_1) \cap (z_2 = 1) \Big] \Bigg)
\end{aligned}
$$

$$
\begin{aligned}
&P(\hat{z}_2 = 1 | z_1 = 1) P(z_1 = 1) + P(\hat{z}_1 = 1 | z_2 = 1) P(z_1 = 1) \\
&= \int_{\{x \in X, \hat{z}_2 = 1\}} f(x; \hat{\theta}_1) dx \hat{\Pi}_1 + \int_{\{x \in X, \hat{z}_1 = 1\}} f(x; \hat{\theta}_2) dx \hat{\Pi}_2
\end{aligned}
$$

3. Analyse dscriminante probabiliste en pratique

On calcule plusieur modèles celui qui minimise l'erreur de classement.

Selon les données auquel'on a fait (stndard | bruitées | grand dimension etc)

On adopte un modèle (Gaussienne, studient,…)

A savoir faire dans un contexte professionel :
- Trouver la libraire à utilisés - interpréter le modèle utilisé avec le doc en ligne - appprter une solution ADP au problème posé

4. Analyse Discriminants Gaussienne

Le context des données sont continus : $(X = R^d)$

On suppose que les loi conditionnlells sont des Gaussienne d-dimiensionelle

$$
f_g(x : \theta_g) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{|\sum_g|}} \exp\Big\{ -\frac{1}{2} (x - \mu g)' \sum_g^{'} (x - \mu g) \Big\}\ o\ x \in R^d
$$

$$
\mu_g \in R^d : le\ centre\ de\ la\ classe\ g
$$

$$
\sum_g \in R^{d \times d}\ SDP : La\ matrice\ de\ covarrance\ de\ la\ claasse\ g
$$

$$
\Theta_g = (\mu_g; \sum_g)
$$

Estimation des Maximum de vraisemblances :

Remark : Dimention $(\mu_g) = d$; dimention $(\sum_g) = \frac{d(d+1)}{2}$

Choix de modèle :

$$\hat{\sum}_g = \begin{cases} \sum_{i=1}^{n} z_{i,g}(X_i - \hat{\mu}_g)(X_i - \hat{\mu}_g)'/n_g \\ hetnoscdastique \ (le \ matrice \ de \ covariance \ sont \ libre) \\ \sum_{g=1}^{G}\sum_{i=1}^{n} z_{i,g}(X_i - \hat{\mu}_g)(X_i - \hat{\mu}_g)(X_i - \hat{\mu}_g)'/n \\ honoscdastique \ (la \ matrices \ sont \ gals) \end{cases}$$

Addition : Matrice de Toeplitz

Exercice : simple à la main

Client	Flux	Solcable
1	3	N
2	4	N
3	5	N
4	5	o
5	6	o
6	7	o
7	9	o
8	5	o

M. Li est un nouveau client, est-il solvable (o) ou non (N) sachant que son flux est 5,2?

Le modèle :

- Solvable : Classe 1
- Non solvable : Classe 2 x = Le flux

$$(x|Classe1) \sim f_1(x;\theta_1) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma_1}\exp\left(\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right) = f_1(X_i;\theta_1) Avec\ \theta_1 = (\mu_1;\sigma_1)$$

$$(x|Classe2) \sim f_1(x;\theta_2) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma_2}\exp\left(\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right) = f_2(X_i;\theta_2) Avec\ \theta_2 = (\mu_2;\sigma_2)$$

$$\pi(Classe1) = \pi_1 \qquad \pi(Classe2) = \pi_2$$

Estimation de maramètre

Classe	Poid	Centre	Variance
g = 1	$\hat{\pi}_1 = 5/8$	$\hat{\mu}_1 = 6.4$	$\hat{\sigma}_1^2 = 2.24$
g = 2	$\hat{\pi}_2 = 3/8$	$\hat{\mu}_2 = 4.0$	$\hat{\sigma}_2^2 = 0.67$

Calculer des scores

$$t_1(L_i) = \frac{\hat{\pi}_1 f_1(5.2; \hat{\sigma}_1)}{\sum_{j=1}^{2} \hat{\pi}_j f_j(5.2; \hat{\sigma}_j)} = 0.66$$

$$t_2(L_i) = \frac{\hat{\pi}_2 f_2(5.2; \hat{\sigma}_2)}{\sum_{j=1}^{2} \hat{\pi}_j f_j(5.2; \hat{\sigma}_j)} = 0.34$$

Décision

$$t_1(L_i) > t_2(L_i)$$

Donc, en affecte M.Li à la classe 1 (solvable).

```
data = data.frame(x = c(3,4,5,5,6,7,9,5), z= c(2,2,2,1,1,1,1,1))
attach(data)
library(Rmixmod)
```

```
## Loading required package: Rcpp

## Rmixmod v. 2.1.8 / URI: www.mixmod.org
```

```
learn <- mixmodLearn(data$x,knownLabels = as.factor(data$z), models = mixmodGaussianModel(listMode
# knowLabels pour lui apprend que c'est d'une flux
learn
```

```
## ****************************************
## *** INPUT:
## ****************************************
## * nbCluster =  2
## * criterion =  BIC
## ****************************************
## *** MIXMOD Models:
## * list =  Gaussian_pk_Lk_Ck
## * This list includes only models with free proportions.
## ****************************************
## * data (limited to a 10x10 matrix) =
## [1] 3 4 5 5 6 7 9 5
```

```
## * ... ...
## * knownLabels =  2 2 2 1 1 1 1 1
##
##
## ****************************************
## *** BEST MODEL OUTPUT:
## *** According to the BIC criterion
## ****************************************
## * nbCluster   =  2
## * model name  =  Gaussian_pk_Lk_Ck
## * criterion   =  BIC(46.5012)
## * likelihood  =  -18.0520
## ****************************************
## *** Cluster 1
## * proportion =  0.6250
## * means      =  6.4000
## * variances  =  2.2400
## *** Cluster 2
## * proportion =  0.3750
## * means      =  4.0000
## * variances  =  0.6667
## ****************************************
## * Classification with MAP:
##             | Cluster 1 | Cluster 2 |
## ----------- ----------- -----------
##  Cluster 1 |         5 |         0 |
##  Cluster 2 |         0 |         3 |
## ----------- ----------- -----------
## * Error rate with MAP =  0.00 %
## ****************************************
```

Voici le doccier, chercher le mot 'BIC' pour trouver la formul

```r
new <-  data.frame(x = c(5.2))
prédiction <- mixmodPredict(data = new, classificationRule = learn['bestResult'])
prédiction
```

```
## ****************************************
## *** INPUT:
## ****************************************
## * nbCluster   =  2
```

```
## * model name  =  Gaussian_pk_Lk_Ck
## * criterion   =  BIC(46.5012)
## * likelihood  =  -18.0520
## ****************************************
## *** Cluster 1
## * proportion =  0.6250
## * means      =  6.4000
## * variances  =  2.2400
## *** Cluster 2
## * proportion =  0.3750
## * means      =  4.0000
## * variances  =  0.6667
## ****************************************
## * Classification with MAP:
##            | Cluster 1 | Cluster 2 |
## ----------- ----------- -----------
##  Cluster 1 |         5 |         0 |
##  Cluster 2 |         0 |         3 |
## ----------- ----------- -----------
## * Error rate with MAP =  0.00 %
## ****************************************
## * data (limited to a 10x10 matrix) =
##   x
## 5.2
## * ... ...
##
##
## ****************************************
## *** PREDICTION:
## ****************************************
## * partition     =  1
## * probabilities =  0.6600 0.3400
## ****************************************
```

Exercice : On considère les données finance de Rmixmid. à quelle classe d'entreprise healthy|banckrucy affectez-vous la première entreprise de 2003 par une méthode d'AD gaussienne bassé sur l'entreprise de 2002.

```r
data("finance")
head(finance)
```

```
##   Year     Health EBITDA.Total.Assets Value.Added.Total.Sales Quick.Ratio
## 1 2002 bankruptcy            -0.00491                 0.21345     0.09041
## 2 2002 bankruptcy             0.08496                 0.11273     0.94598
## 3 2002    healthy             0.45284                 0.48414     1.37340
## 4 2002    healthy             0.20980                 0.39530     1.27090
## 5 2002 bankruptcy            -0.07732                 0.29466     0.69698
## 6 2002 bankruptcy             0.02914                 0.38550     0.64299
##   Accounts.Payable.Total.Sales
## 1                      0.29409
## 2                      0.28540
## 3                      0.05980
## 4                      0.25352
## 5                      0.05581
## 6                      0.10572
```

```r
unique(finance$Year)
```

```
## [1] 2002 2003
## Levels: 2002 2003
```

```r
train = finance[finance$Year == '2002',3:6]
ztrain = finance[finance$Year == '2002', 2]
head(train)
```

```
##   EBITDA.Total.Assets Value.Added.Total.Sales Quick.Ratio
## 1            -0.00491                 0.21345     0.09041
## 2             0.08496                 0.11273     0.94598
## 3             0.45284                 0.48414     1.37340
## 4             0.20980                 0.39530     1.27090
## 5            -0.07732                 0.29466     0.69698
## 6             0.02914                 0.38550     0.64299
##   Accounts.Payable.Total.Sales
## 1                      0.29409
## 2                      0.28540
## 3                      0.05980
## 4                      0.25352
## 5                      0.05581
## 6                      0.10572
```

```r
head(ztrain)
```

```
## [1] bankruptcy bankruptcy healthy    healthy    bankruptcy bankruptcy
## Levels: bankruptcy healthy
```

```r
learn <- mixmodLearn(train,knownLabels = as.factor(ztrain), models = mixmodGaussianModel(listModel
learn
```

```
## ****************************************
## *** INPUT:
## ****************************************
## * nbCluster =  2
## * criterion =  BIC
## ****************************************
## *** MIXMOD Models:
## * list =  Gaussian_pk_Lk_Ck
## * This list includes only models with free proportions.
## ****************************************
## * data (limited to a 10x10 matrix) =
##    EBITDA.Total.Assets Value.Added.Total.Sales Quick.Ratio
## 1  -0.00491            0.2135                  0.09041
## 2  0.08496             0.1127                  0.946
## 3  0.4528              0.4841                  1.373
## 4  0.2098              0.3953                  1.271
## 5  -0.07732            0.2947                  0.697
## 6  0.02914             0.3855                  0.643
## 7  1e-05               0.1955                  0.6889
## 8  0.5608              0.4002                  1.64
## 9  -0.02126            0.1665                  0.1583
## 10 -0.00937            0.1521                  0.6395
##    Accounts.Payable.Total.Sales
## 1  0.2941
## 2  0.2854
## 3  0.0598
## 4  0.2535
## 5  0.05581
## 6  0.1057
## 7  0.24
## 8  0.09743
## 9  0.2836
## 10 0.1789
## * ... ...
## * knownLabels =  1 1 2 2 1 1 1 2 1 1  ...
##
##
```

```
## ****************************************
## *** BEST MODEL OUTPUT:
## *** According to the BIC criterion
## ****************************************
## * nbCluster   =  2
## * model name  =  Gaussian_pk_Lk_Ck
## * criterion   =  BIC(-792.0487)
## * likelihood  =  483.8817
## ****************************************
## *** Cluster 1
## * proportion =  0.4953
## * means      =  -0.0386 0.2069 0.6089 0.1774
## * variances  = |     0.0298     0.0068     0.0116    -0.0025 |
##                |     0.0068     0.0145     0.0045    -0.0014 |
##                |     0.0116     0.0045     0.1680    -0.0085 |
##                |    -0.0025    -0.0014    -0.0085     0.0090 |
## *** Cluster 2
## * proportion =  0.5047
## * means      =  0.1662 0.2749 1.0661 0.1079
## * variances  = |     0.0118     0.0045     0.0194    -0.0016 |
##                |     0.0045     0.0142     0.0082     0.0001 |
##                |     0.0194     0.0082     0.2859    -0.0080 |
##                |    -0.0016     0.0001    -0.0080     0.0052 |
## ****************************************
## * Classification with MAP:
##            | Cluster 1 | Cluster 2 |
## ----------- ----------- -----------
##  Cluster 1 |       212 |         0 |
##  Cluster 2 |         0 |       216 |
## ----------- ----------- -----------
## * Error rate with MAP =  0.00 %
## ****************************************
```

```r
nrow(train)
```

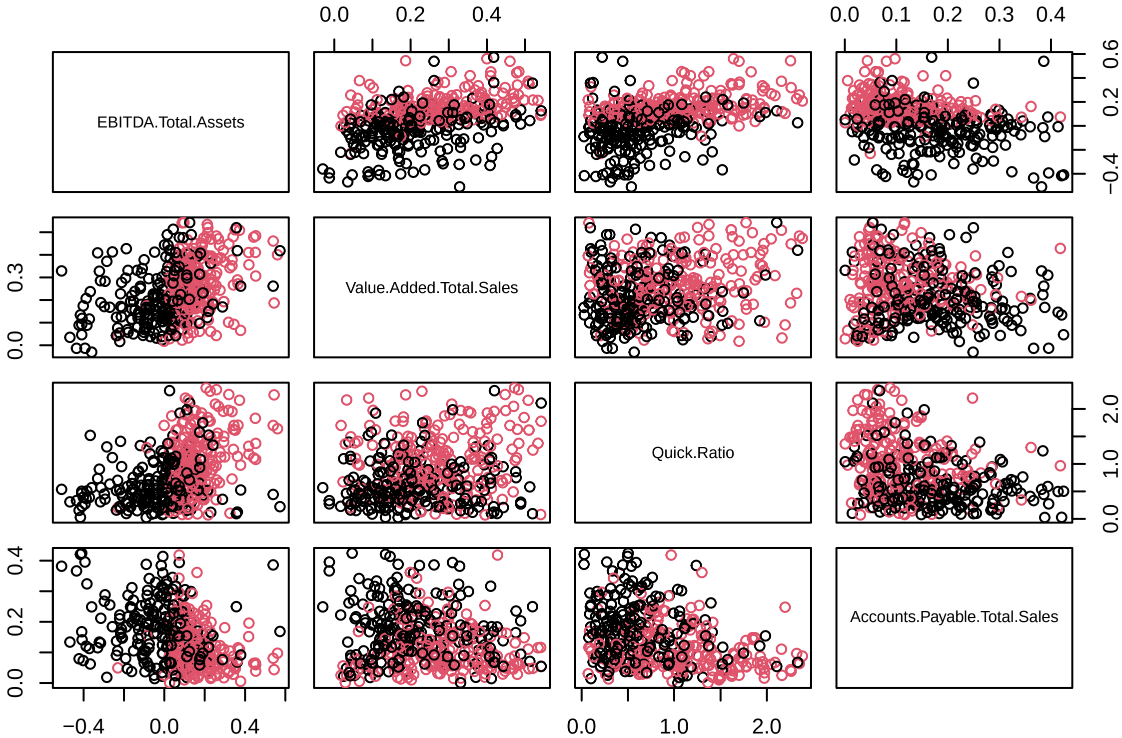
```
## [1] 428
```

```r
sum(ztrain == 'bankruptcy')/428
```

```
## [1] 0.4953271
```

```r
# Pour retrouver le codage de classes en comparant les proportion mais on ne pourra le faire à cha
```

```r
plot(train, col = as.factor(ztrain))
```




```r
new <- finance[finance$Year =='2003', 3:6][1,]
prédiction <- mixmodPredict(data = new, classificationRule = learn['bestResult'])
prédiction
```

```
## ****************************************
## *** INPUT:
## ****************************************
## * nbCluster   =  2
## * model name  =  Gaussian_pk_Lk_Ck
## * criterion   =  BIC(-792.0487)
## * likelihood  =  483.8817
## ****************************************
## *** Cluster 1
## * proportion =  0.4953
## * means      =  -0.0386 0.2069 0.6089 0.1774
## * variances  = |     0.0298     0.0068     0.0116    -0.0025 |
##                |     0.0068     0.0145     0.0045    -0.0014 |
##                |     0.0116     0.0045     0.1680    -0.0085 |
##                |    -0.0025    -0.0014    -0.0085     0.0090 |
```

```
## *** Cluster 2
## * proportion =  0.5047
## * means      =  0.1662 0.2749 1.0661 0.1079
## * variances  = |     0.0118     0.0045     0.0194    -0.0016 |
##              |     0.0045     0.0142     0.0082     0.0001 |
##              |     0.0194     0.0082     0.2859    -0.0080 |
##              |    -0.0016     0.0001    -0.0080     0.0052 |
## ****************************************
## * Classification with MAP:
##            | Cluster 1 | Cluster 2 |
## ----------- ----------- -----------
##  Cluster 1 |       212 |         0 |
##  Cluster 2 |         0 |       216 |
## ----------- ----------- -----------
## * Error rate with MAP =  0.00 %
## ****************************************
## * data (limited to a 10x10 matrix) =
##          EBITDA.Total.Assets      Value.Added.Total.Sales
##                       0.1029                       0.2388
##                  Quick.Ratio Accounts.Payable.Total.Sales
##                       0.8886                       0.2412
## * ... ...
##
##
## ****************************************
## *** PREDICTION:
## ****************************************
## * partition     =  1
## * probabilities =  0.5725 0.4275
## ****************************************
```

Elle est dans le classe 1, $t_1 = 0.5725$ et $t_2 = 0.4275$

Question subsidiaine L'AD Conssionnne hétéroscélastique (Gaussin_pk_Lk_Ck) affect la 1ière ebtreeprise de 2003 à la classe bankrucptcy avec probabilité 0.5725

Que dire de la classe de cette entrerprise en AD Gaussienne homoscédastique (Gaussinne_pk_Lk_Ck) Lequel des deux modèle homoscélasticique /hétéscélasticisituque le créteur BIC préfère-t-il?

	Proba Bankcuryty healty	BIC
Gaussien Homoscédastique	5.5476 vs 0.4524	760/2

	Proba Bankcuryty healty	BIC
Gaussien Hétéscédastique	0.5725 vs 0.4275	792/2

Si on droit choisir un le plus grand, c'est à dire le deuxième

21092023 8. Reregression logistique (RL)

On suppose le observation répation en deux classe et les descriptoon condition la RL peut être à plus de deux classe : Voir CH2 REF 6 p196

Objectif : obtenir une fonction discriminante somple (linéaire) avec un modèle semi paramétrique (économie en paremets)

On reprend le nature de 2

Etant donné $x = (x1; ...; x_d)' \in R^d$ on souhaite estimer la classe $z = (z_1, z_2) \in 0, 1^2$ de x

Le modème : $p(z_1 = 1|x) = \frac{e^{\beta' x + \alpha}}{1+e^{\beta' x+\alpha}}$ avec $\beta = (1; ...; \beta d)' \in R^d \alpha \in R$ Le modèle : S(x) = 'x +

La regression de classement

$$\begin{aligned} \hat{z}_1 = 1 \leftrightarrow S(x) > 0 \quad &\leftrightarrow P(Z_1 = 1) > 1/2 \\ \hat{z}_2 = 1 \leftrightarrow S(x) \leq 0 \quad &\leftrightarrow P(Z_2 = 1) \geq 1/2 \end{aligned}$$

Justification :

$$P(Z_1 = 1|x) > 1/2 \leftrightarrow \frac{e^{s(x)}}{1+e^{s/x_1}} > \frac{1}{2}$$

$$\begin{aligned} 2e^{s(x)} &> 1 + e^{s/x_1} \\ e^{s(x)} &> 1 \\ S(x) &> 0 \end{aligned}$$

La fonction discriminante $\{n \in R^d ; S(x) = 0\}$ effet confiance

Estimation du paramétrique : $\theta = (\alpha, \beta)' \in R^{d+1}$

Vraisemblance de θ :

$$\begin{aligned} P(\theta; \{(x; z_i); i : 1, ..., n\}) &= \Pi_{i=1}^{n} [p(z_{i,1} = \frac{1}{x_i})]^{z_{i,1}} [p(z_{i,2} = \frac{1}{x_i})]^{z_{i,2}} \\ &= \Pi_{i=1}^{n} \left[\frac{\exp(\beta' x_i + \alpha)}{1 + \exp(\beta' X_i + \alpha)} \right]^{z_{i,1}} \left[\frac{1}{1 + \exp(\beta' X_i + \alpha)} \right]^{z_{i,2}} \\ &= \Pi_{i=1}^{n} \frac{\exp(\beta' x_i + \alpha)^{z_{i,1}}}{1 + \exp(\beta' X_i + \alpha)} \end{aligned}$$

La log vraisemblance de δ

$$P = (\theta, \{1; z_i\}; i = 1, ..., n) = \sum_{i=1}^{n} \Big\{ z_{i,1}(\beta' x_i + \alpha) - \log\left(1 + \exp(\beta' x_i + \alpha)\right) \Big\}$$

Equation de vraisemblance

$$\forall (\beta, \alpha), P(\theta; \{(x; z_i); i : 1, ..., n\}) = O_{R^{d+1}}$$

En pratique, on ne sait pas résoudre (*) à la main

Résolution de équation de vraisemblance : - en utilisation de fonction glm de R (voir exemple souvant) - en utilisant une fonction d'ompimisation plus générale (option de R)

Un exemple :

Salaire	âge	classe
1,7	28	A
2,2	38	B
3,4	43	B
3,5	54	A

A quelle classe M. martin affect t elle? M. martin: slaire = 2,1; âge 34.

```r
# donnees
# http://alexandrelourme.free.fr/M2IREF/SCORING/LRscript
train=data.frame(sal=c(17,22,34,35)/10,age=c(28,38,43,54),classe=c('A','B','B','A')) ; attach(trai
test=data.frame(sal=2.1,age=34) # test data
```

```r
# plots

plot(sal,age,cex=2,col=as.factor(classe),pch=as.character(classe)) # train data
points(test,pch=19,cex=2,col='green') # test data

# Logistic Regression with glm

rule=glm(as.factor(classe)~sal+age,family=binomial(link='logit')) # model parameter inference
rule$coefficients # estimation of alpha = rule$coefficients[1] ; estimation of beta_1 = rule$coeff
```

```
## (Intercept)         sal         age
##   1.1576023   2.6120143  -0.2005145
```

```r
abline(a=-rule$coefficients[1]/rule$coefficients[3],b=-rule$coefficients[2]/rule$coefficients[3])

#a = intercept
#b = la pente

score <- predict(rule,new=test) ; print(score) # the value of :  beta_1*sal + beta_2*age + alpha c
```

```
##          1
## -0.1746618
```

```r
prob <- exp(score)/(1+exp(score)) ; print(prob) # probability of belonging to Class 1, 属于 classe
```

```
##         1
## 0.4564452
```

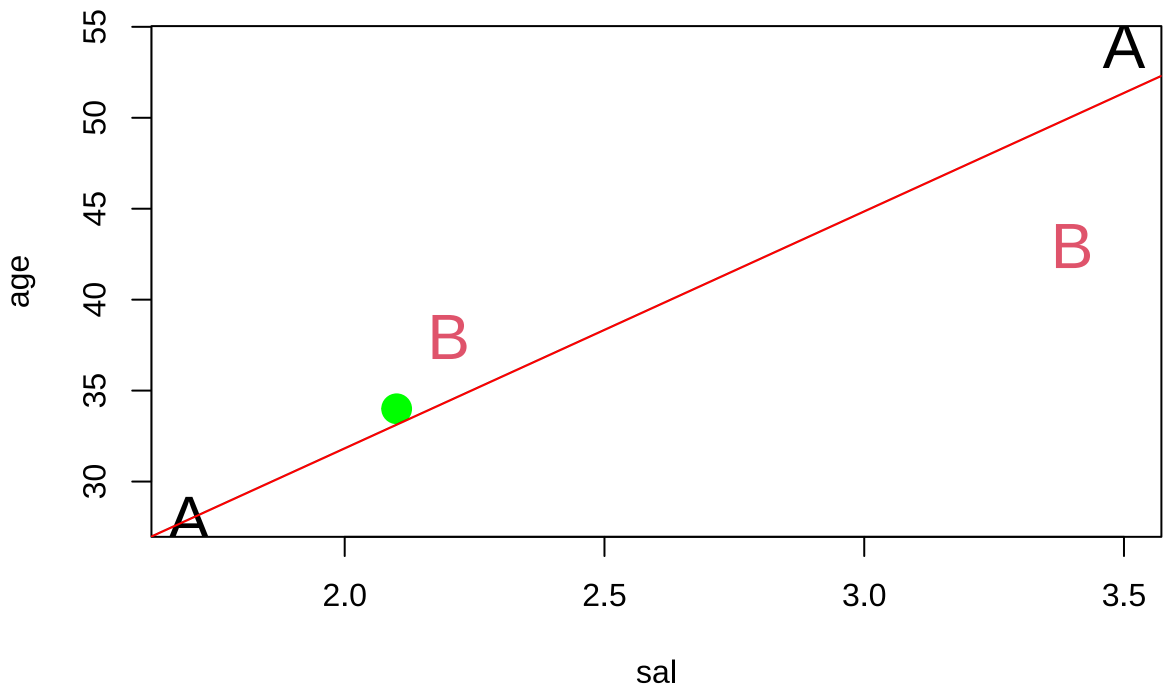
```r
predict(rule,new=data.frame(sal=sal,age=age)) # the value of  : beta_1*sal + beta_2*age + alpha co
```

```
##           1           2           3           4
## -0.01638029 -0.71551848  1.41632601 -0.52813244
```

```r
# On préfère l'érreur de classe plus miminum

# Logistic Regression with optim

mll <- function(data,para){out=0
for (i in 1:4){sci=para[1]*data[i,1]+para[2]*data[i,2]+para[3] # para[1]=beta_1 (sal) ; para[2]=be
out=out+sci*(data[i,3]=='B')-log(1+exp(sci))}
return(-out)}  #  maximizing the log-likelihood <=> minimising the opposite of the log-likelihood

res <- optim(par=c(3,1,2),fn=mll,data=train) # the LR parameters are within res$par
res$par # res$par[1]=alpha_1 ; res$par[1]=beta_1 ;  res$par[2]=beta_2 ; res$par[3]=alpha
```

```
## [1]  2.6147115 -0.2006959  1.1569023
```

```r
abline(a=-res$par[3]/res$par[2],b=-res$par[1]/res$par[2],col='red') # the discriminant rule
```




```
prob <- function(data){sal=data[1]; age=data[2];lin=res$par[1]*sal+res$par[2]*age+res$par[3];exp(l

prob(test) # probability for the test data to belong to class 1
```

```
##          sal
## 1 0.4561473
```

```
prob(train) # probabilities that train data belong to Class 1
```

```
##          sal
## 1 0.4956071
## 2 0.3280154
## 3 0.8048676
## 4 0.3707073
```

$$S(M.Martin) = \hat{\alpha} + \hat{\beta}_1 \times 2,1 + \hat{\beta}_2 \times 34$$

Comment savoir si

$$[\underbrace{\hat{1} \leftrightarrow \hat{A} \ et \ \hat{2} \leftrightarrow B}_{Choix\ 1}] \ ou \ [\underbrace{\hat{1} \leftrightarrow B \ et \ \hat{2} \leftrightarrow A}_{Choix\ 2}]$$

Salaire	Âge	Classe estimée	Classe estimée
1,7	28	$\hat{B}$	$\hat{A}$

Salaire	Âge	Classe estimée	Classe estimée
2,2	38	$\hat{B}$	$\hat{A}$
3,4	43	$\hat{A}$	$\hat{B}$
3,5	54	$\hat{B}$	$\hat{A}$
		Erreur de classement	Erreur de classement
		75%	25%

$$S(x) = 0\alpha + R_1(salaire) + R_2(\hat{a}ge) = 0\hat{a}ge = -\frac{\beta_1}{\beta_2}Salaire = -\frac{\alpha}{\beta}$$

M. Martin est estimer dans la classe A

Contrôle Scoring prochaine séance, Chapitr 1-2 , 1h30, Judi prochaine 13h15, 取代 actuariat ，可以使用电脑

Projet 2