

Corporate Risk - Scores & Rules with R

Let us consider the data `finance` available in `Rmixmod` package.

Gaussian model-based Bayesian scores

1. Load the data into a R session.

```
library(Rmixmod)
data(finance)
attach(finance)
```

We consider the financial ratios : `EBITDA.Total.Assets`, `Value.Added.Total.Sales`, `Quick.Ratio`, `Accounts.Payable.Total.Sales` as predictors of the health of the firms.

2. Estimate the parameters of the Gaussian classes.

```
library(Rmixmod)
learn <- mixmodLearn(finance[Year=='2002',3:6],knownLabels=finance[Year=='2002',2])
```

3. Compare the ratios of healthy/bankruptcy firms.

```
learn

## *****
## *** INPUT:
## *****
## * nbCluster = 2
## * criterion = CV
## *****
## *** MIXMOD Models:
## * list = Gaussian_pk_Lk_C
## * 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 CV criterion
## *****
## * nbCluster = 2
## * model name = Gaussian_pk_Lk_C
## * criterion = (0.1822)
## * likelihood = 444.9579
## *****
## *** Cluster 1
## * proportion = 0.4953
## * means = -0.0386 0.2069 0.6089 0.1774
## * variances = | 0.0226 0.0064 0.0186 -0.0023 |
## | 0.0064 0.0166 0.0076 -0.0006 |
## | 0.0186 0.0076 0.2728 -0.0095 |
## | -0.0023 -0.0006 -0.0095 0.0079 |
## *** Cluster 2
## * proportion = 0.5047
## * means = 0.1662 0.2749 1.0661 0.1079
## * variances = | 0.0172 0.0049 0.0142 -0.0017 |
## | 0.0049 0.0126 0.0058 -0.0005 |
## | 0.0142 0.0058 0.2076 -0.0073 |
## | -0.0017 -0.0005 -0.0073 0.0060 |
## *****
## * Classification with MAP:
## | Cluster 1 | Cluster 2 |
## -----
## Cluster 1 | 212 | 0 |
## Cluster 2 | 0 | 216 |
## -----
## * Error rate with MAP = 0.00 %
## *****
```

4. Compare the average value of Value.Added.Total.Sales for healthy/bankruptcy firms.
5. Compare the variance of Quick Ratio for healthy/bankruptcy firms.
6. Predict the health of firms in 2003 thanks to the parameters learnt in 2002.

```
prediction <- mixmodPredict(data=finance[Year=='2003',3:6],
classificationRule=learn["bestResult"])
```

One considers as a score the conditional probability to belong to the healthy class and one firm is estimated as healthy if the score is greater than $\tau = 0.5$.

7. What is the score of ADTech, the second firm of Year 2003?

```
ADTech=prediction[6][2,]
ADTech
## [1] 0.8125288 0.1874712
```

8. Would you say ADTech is healthy or not?
9. Estimate the class of all firms in 2003?

```
estimated_class <- prediction[5]
```

10. What is the misclassification error rate?

```
estimated_class <- prediction[5]
true_class <- Health[Year=='2003']
table(true_class,estimated_class)

##           estimated_class
## true_class      1      2
##  bankruptcy 162    58
##   healthy   54   187
```

11. How many firms are wrongly considered as healthy?
12. How many firms are wrongly considered as non healthy?
13. Compute and interpret the false positive rate.
14. Compute and interpret the true positive rate.
15. Interpret the probabilities within: `prediction[6][,2]`.
16. Interpret the output of: `prediction[6][,2]>0.5`.

One changes the allocation threshold: a firm is estimated as healthy if the score is greater than $\tau = 0.4$.

17. Would you still say that ADTech is healthy or not?
18. Estimate the class of all firms in 2003.

```
estimated_class <- prediction[6][,2]>0.4
```

19. What is the misclassification error rate?

```
table(true_class,estimated_class)

##           estimated_class
## true_class  FALSE  TRUE
##  bankruptcy   141    79
##   healthy     38   203
```

20. Update the answers to Questions 11 to 14.
