Kubra Iqbal

Homework 4.

2) 20 points): A common application of Discriminant Analysis is the classification of bonds into various bond rating classes. These ratings are intended to reflect the risk of the bond and influence the cost of borrowing for companies that issue bonds. Various financial ratios culled from annual reports are often used to help determine a company's bond rating. The Excel spreadsheet BondRating.xls (XLS) contains two sheets named Training data and Validation data. These are data from a sample of 95 companies selected from COMPUSTAT financial data tapes. The company bonds have been classified by Moody's Bond Ratings (1980) into seven classes of risk ranging from AAA, the safest, to C, the most risky. The data include ten financial variables for each company.

These are:

LOPMAR: Logarithm of the operating margin,

LFIXMAR: Logarithm of the pretax fixed charge coverage,

LTDCAP: Long-term debt to capitalization,

LGERRAT: Logarithm of total long-term debt to total equity,

LLEVER: Logarithm of the leverage,

LCASHLTD: Logarithm of the cash flow to long-term debt,

LACIDRAT: Logarithm of the acid test ratio,

LCURRAT: Logarithm of the current assets to current liabilities,

LRECTURN: Logarithm of the receivable turnover,

LASSLTD: Logarithm of the net tangible assets to long-term debt.

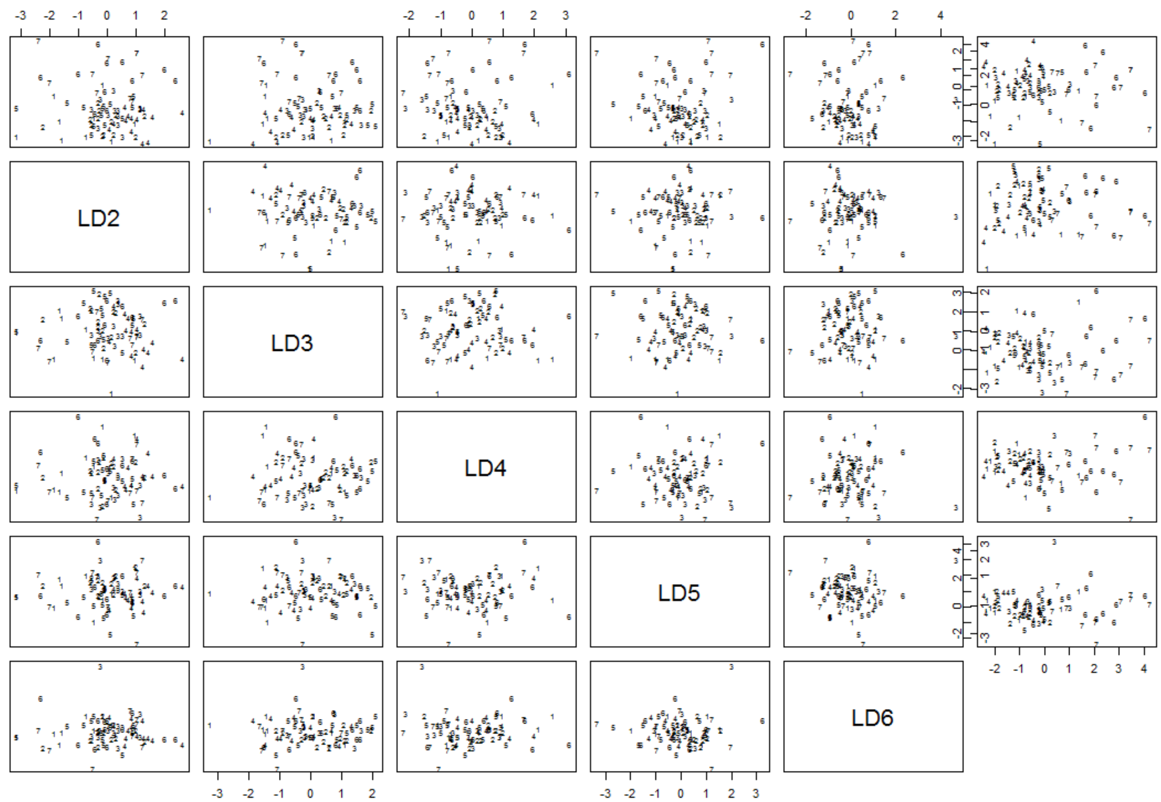
The data are divided into 81 observations in the Training data sheet and 14 observations in the Validation data sheet. The bond ratings have been coded into numbers in the column with the title CODERTG, with AAA coded as 1, AA as 2, etc. Develop a Linear Discriminant Analysis model to classify the bonds in the Validation data sheet.

a) What is the performance of the classifier on the training data? Notice that there is order in the class variables (i.e., AAA is better than AA, which is better than A,…).

b) What is the performance of the classifier on the validation data?

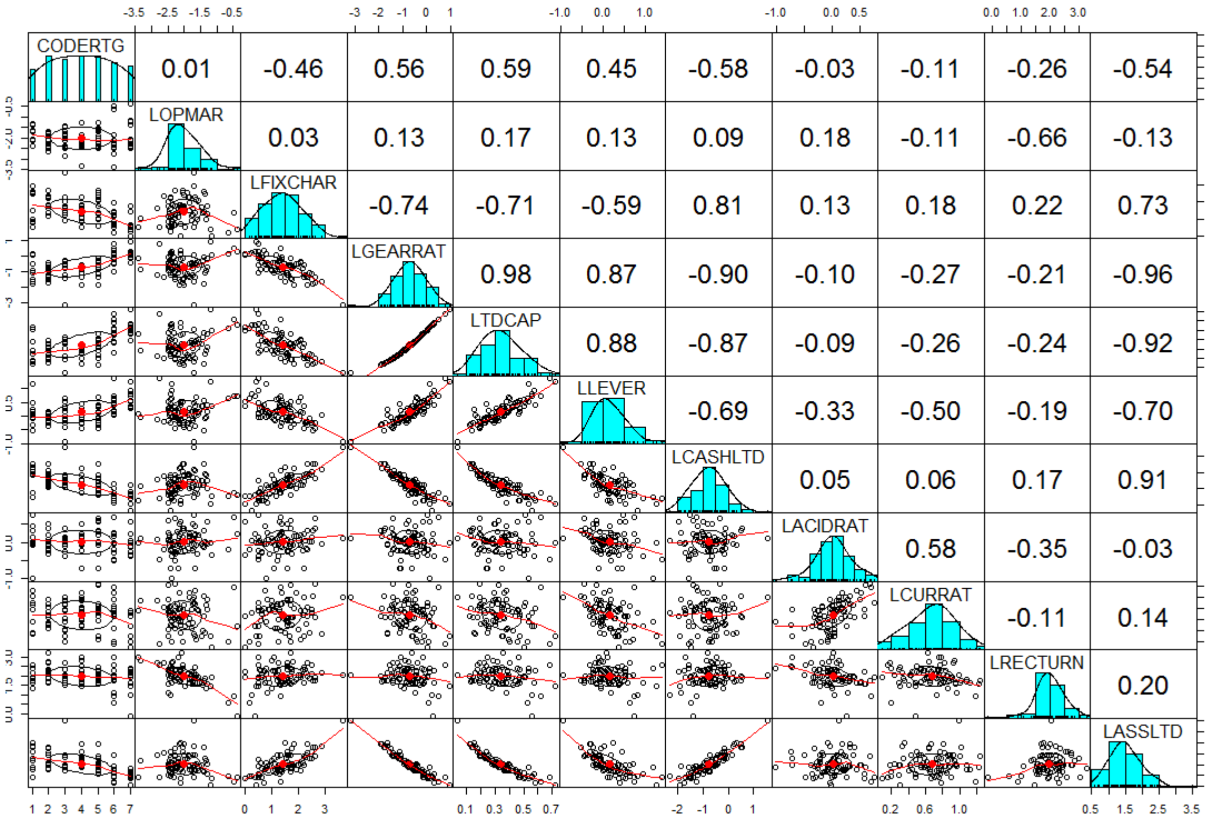
c) Would certain misclassification errors be worse than others? If so, how would you suggest measuring this?

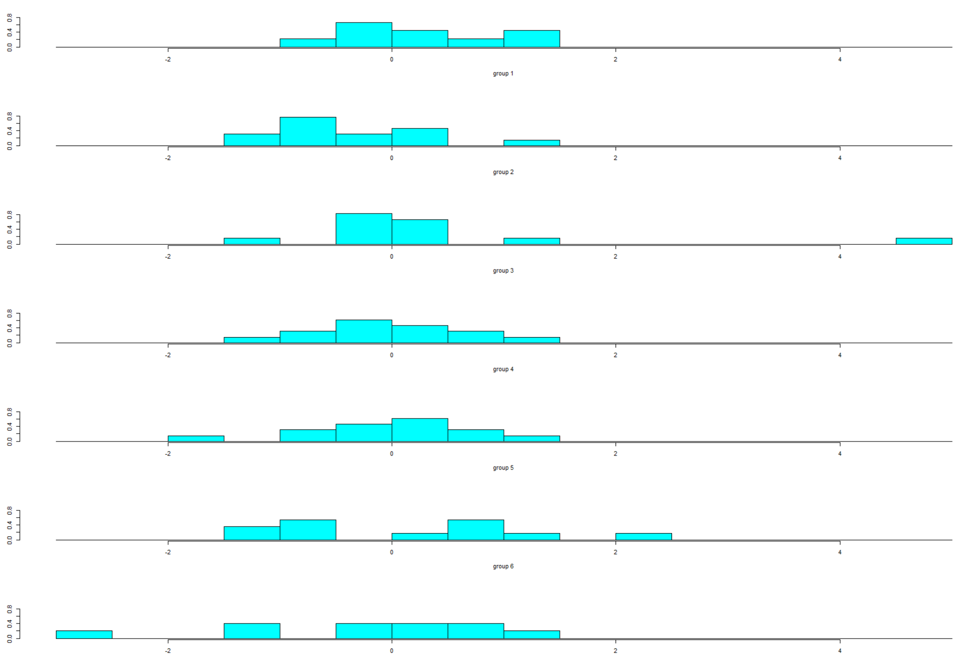
```
> #comparing results
> table(p, newdata$CODERTG)

p    1  2  3  4  5  6  7
  1  4  1  0  0  1  1  0
  2  3  7  3  1  1  0  0
  3  0  1  6  0  1  0  2
  4  1  2  2 11  2  0  1
  5  0  2  1  1  8  1  0
  6  1  0  0  0  0  8  1
  7  0  0  0  0  0  1  6
>
```



```
Group means:
     LOPMAR  LFIXCHAR    LGEARRAT    LTDCAP      LLEVER   LCASHLTD     LACIDRAT   LCURRAT LRECTURN  LASSLTD
1 -1.738889 1.6637778 -0.99555556 0.2881111  0.12388889 -0.3940000  0.059888889 0.6932222 1.943889 1.804000
2 -2.094385 1.8042308 -1.05315385 0.2641538 -0.08338462 -0.3925385 -0.003692308 0.6640769 2.266308 1.733462
3 -2.017917 1.7306667 -0.94075000 0.3034167  0.04291667 -0.4003333  0.017500000 0.6387500 2.074250 1.693417
4 -2.213923 1.3204615 -1.01200000 0.2704615 -0.02153846 -0.5720769 -0.063230769 0.7600769 2.032077 1.721769
5 -1.981846 1.7073077 -0.75800000 0.3272308  0.07430769 -0.7765385  0.137076923 0.7471538 1.950000 1.510077
6 -2.078545 0.9529091 -0.07790909 0.4812727  0.44972727 -1.4103636 -0.033181818 0.7031818 1.818182 1.103182
7 -1.783600 0.5873000  0.10860000 0.5248000  0.64370000 -1.4720000 -0.031600000 0.4642000 1.650000 0.993700

Coefficients of linear discriminants:
                LD1        LD2        LD3          LD4           LD5        LD6
LOPMAR   -0.7720156  -2.993776 -1.0902999   1.19056396   0.003079991 -1.0907388
LFIXCHAR  0.3309649  -1.032219  2.0342609  -0.17225468  -0.566130362  0.4446614
LGEARRAT  2.0228900 -13.206606  4.3603205  30.56370258  19.296973115 -8.6572293
LTDCAP   27.6725970  15.434851  1.0663233 -30.15183168   0.636947862 22.5703473
LLEVER   -5.2113899   4.540020 -5.2197916 -13.97013291 -12.485287860  4.5123115
LCASHLTD -0.8040312   3.684976 -0.6103313  -1.47884309   2.343115368  2.1285439
LACIDRAT -0.2978150  -3.360777 -0.7014467  -0.09884748   0.507853522 -0.9383520
LCURRAT  -2.0007312   2.040593 -1.1419790   1.51718949  -2.677213623  3.2930473
LRECTURN -1.1369903  -2.245231 -0.6432160   0.81809242   0.686713979 -0.9182123
LASSLTD   5.2328461 -14.461158  1.3481935  26.33072526  16.502239043 -5.7011832

Proportion of trace:
   LD1    LD2    LD3    LD4    LD5    LD6
0.6309 0.1209 0.1005 0.0705 0.0587 0.0186
```

```
> # Compare the results of the prediction
> table(p, newdata$CODERTG)

p    1  2  3  4  5  6  7
  1  4  1  0  0  1  1  0
  2  3  7  3  1  1  0  0
  3  0  1  6  0  1  0  2
  4  1  2  2 11  2  0  1
  5  0  2  1  1  8  1  0
  6  1  0  0  0  0  8  1
  7  0  0  0  0  0  1  6
>
> # Setting "CV = T" will have the lda function perform
> # "Leave-one-out" cross-validation
> LDA=lda(CODERTG ~ ., data=newdata, CV=T)
> table(LDA$class, BondRating$CODERTG)

    1 2 3 4 5 6 7
  1 0 2 0 1 2 2 1
  2 3 4 3 2 1 0 0
  3 0 3 3 0 1 2 2
  4 3 2 2 8 3 0 1
  5 1 2 2 2 6 1 0
  6 1 0 0 0 0 4 4
  7 1 0 2 0 0 2 2
> coef(LDA)
NULL
>
```


CODERTG 0.01 -0.46 0.56 0.59 0.45 -0.58 -0.03 -0.11 -0.26 -0.54
LOPMAR 0.03 0.13 0.17 0.13 0.09 0.18 -0.11 -0.66 -0.13
LFIXCHAR -0.74 -0.71 -0.59 0.81 0.13 0.18 0.22 0.73
LGEARRAT 0.98 0.87 -0.90 -0.10 -0.27 -0.21 -0.96
LTDCAP 0.88 -0.87 -0.09 -0.26 -0.24 -0.92
LLEVER -0.69 -0.33 -0.50 -0.19 -0.70
LCASHLTD 0.05 0.06 0.17 0.91
LACIDRAT 0.58 -0.35 -0.03
LCURRAT -0.11 0.14
LRECTURN 0.20
LASSLTD


group 1
group 2
group 3
group 4
group 5
group 6
group 7

Yes.

The prediction accuracy is not same for all the scenarios. To improve these results, I think there should be an accuracy included for each segment not just to the particular performance. Another variable should also be added to predict the error.