Financial Analytics

Group work/preentations

Download the file „FA – Group_ data.csv".

The dataset is about bankruptcy prediction of Polish companies. The data was collected from Emerging Markets Information Service, which is a database containing information on emerging markets around the world. The data comes from the years 2000-2013. The data contains financial information from 1st year of the forecasting period and corresponding class label that indicates bankruptcy status after 5 years.

The variable definitions can be found on the next page.

Question 1:

The Altman z-score is a formula for predicting bankruptcies. It contains of five variables:
- working capital / total assets
- retained earnings / total assets
- EBITDA / total assets
- book value of equity / total liabilities
- sales / total assets

Create a data frame with only these five variables and the bankruptcy outcome.
Clean the data for missing data.

A) Create a table with the means and standard deviations of all variables.
B) Produce a scatterplot matrix which includes all of the variables in the data set.
C) Perform a logistic regression with class as the response variables and the five other variables as the predictor. What do you observe?
D) Compute the confusion matrix and overall fraction of correct predictions. Explain.
E) Using the validation set approach, estimate the test error of this model. Split the sample into 50/50 training and test data. Compute the validation set error, which is the fraction of the observations in the validation set that are misclassified. What do you observe?
F) Repeat part E using KNN with k=1. What do you observe?
G) Repeat part E using KNN with k=10. What do you observe?

Question 2:

A) Try to improve the prediction of the logistic regression by including additional variables. Explain the reasons for your selected variables in detail. How well are you able to predict a bankruptcy?
B) Extend your analysis to a LASSO approach. First, use two arbitrary lambdas to investigate the behavior of your model. Comment on your findings
C) Second, tune the optimal lambda parameter. Comment on your finding.

Variable definitions:

X1 net profit / total assets

X2 total liabilities / total assets

X3 working capital / total assets

X4 current assets / short-term liabilities

X5 [(cash + short-term securities + receivables - short-term liabilities) / (operating expenses - depreciation)] * 365

X6 retained earnings / total assets

X7 EBIT / total assets

X8 book value of equity / total liabilities

X9 sales / total assets

X10 equity / total assets

X11 (gross profit + extraordinary items + financial expenses) / total assets

X12 gross profit / short-term liabilities

X13 (gross profit + depreciation) / sales

X14 (gross profit + interest) / total assets

X15 (total liabilities * 365) / (gross profit + depreciation)

X16 (gross profit + depreciation) / total liabilities

X17 total assets / total liabilities

X18 gross profit / total assets

X19 gross profit / sales

X20 (inventory * 365) / sales

X21 sales (n) / sales (n-1)

X22 profit on operating activities / total assets

X23 net profit / sales

X24 gross profit (in 3 years) / total assets

X25 (equity - share capital) / total assets

X26 (net profit + depreciation) / total liabilities

X27 profit on operating activities / financial expenses

X28 working capital / fixed assets

X29 logarithm of total assets

X30 (total liabilities - cash) / sales

X31 (gross profit + interest) / sales

X32 (current liabilities * 365) / cost of products sold

X33 operating expenses / short-term liabilities

X34 operating expenses / total liabilities

X35 profit on sales / total assets

X36 total sales / total assets

X37 (current assets - inventories) / long-term liabilities

X38 constant capital / total assets

X39 profit on sales / sales

X40 (current assets - inventory - receivables) / short-term liabilities

X41 total liabilities / ((profit on operating activities + depreciation) * (12/365))

X42 profit on operating activities / sales

X43 rotation receivables + inventory turnover in days

X44 (receivables * 365) / sales

X45 net profit / inventory

X46 (current assets - inventory) / short-term liabilities

X47 (inventory * 365) / cost of products sold
X48 EBITDA / total assets
X49 EBITDA / sales
X50 current assets / total liabilities
X51 short-term liabilities / total assets
X52 (short-term liabilities * 365) / cost of products sold)
X53 equity / fixed assets
X54 constant capital / fixed assets
X55 working capital
X56 (sales - cost of products sold) / sales
X57 (current assets - inventory - short-term liabilities) / (sales - gross profit - depreciation)
X58 total costs /total sales
X59 long-term liabilities / equity
X60 sales / inventory
X61 sales / receivables
X62 (short-term liabilities *365) / sales
X63 sales / short-term liabilities
X64 sales / fixed assets
class dummy variable indicating a bankruptcy (=1)