PS4: EC48E: Fin. Appl. of Machine Learning

Burak Saltoğlu

Instructions

- This is a group project, each group can have up to 3 students.
- Submit your work to Moodle.
- Present your work clearly without code history in your report as in previous assignments.
- For any further questions, you can consult the TA.
- Deadline: **18.05.2023 23.59**

Part A (Classification problem and logistic regression)

You are given a banks corporate loan default statistcs with 1 defaults and 0 survivor companies. Features that are used to explain this defaults are given as; Working Capital (WC), Retained Earnings (RE), Earning before interest and taxes (EBIT), Market value of equity (ME), Sales divided by total liabilities (S). All variables are divided by total assets. You are given the data in the **credit.csv** file.

a. Run a logistic regression model and briefly comment on both the signs and magnitude of the estimated coefficients.

b. Calculate the R^2 of the logistic regression and discuss why it is not a proper measure for classification problem.

c. By assuming your probability of default benchmark is 50%, evaluate your model's performance by filling the following truth table.

True Positive	False Negative
False Positive	True Negative

d. What can be said about the performance of the model by using the above Truth table?

e. Can you calculate the ROC curve statistic for this model?

f. By leaving the last 10% of the sample of the data as a test set, calculate the truth table and ROC curve statistic.

g. Calculate the ROC statistics of the above model by using LOOCV method and k-fold CV where k=10.

h. By using "Best Subset Regression" method as we discussed in the class, which variable combinations out of these 5 variables can be chosen as the best model?

i. Try to improve the predictive power of your model with some nonlinear transformations on the variables (i.e. log or sin or whatever is appropriate).

Part B (Lasso, Ridge and Elastic Net Regression)

By using the data set given in Part A, fit the Ridge, LASSO and Elastic Net regressions by choosing the tuning parameter through Cross Validation method (in R you can use *glmnet* library, but you are free to choose any other R or Pyhton library)

a. Comment on each of the coefficents.

b. Calculate the truth table statistics for the k-fold cross validation method for k=10.

c. Calculate ROC curve for each of the 3 models by using only in sample methods.

d. After fixing the tuning parameter, use cross validation method to choose which of the 3 regularization method fare best in terms of ROC.

Part C (Unsupervised learning)

As we did in the lectures please redo the PCA on the Turkish Treasury Yield Curve.

a. Make a descriptive analysis on the Turkish local interest rates as for levels and differences (r(t)-r(t-1)) (i.e. mean, median, skewness, kurtosis)

b. Calculate how many months (out of 111) 10 year and 2 year yields are positive.

c. Find at least three periods where the yield curve is flattened. Describe what type of bond trading strategies can be developed for these periods.

d. Find some periods where the yield curve steepens. What does this imply in terms of bond trading?

e. Conduct the PC Analysis on the yield curve for 2013, 2018, 2019 seperately. Do you notice any structural difference in each year?

Part D Neural Net

Use Logit1.csv to create a neural net model. Split %20 of the data to test the models. Construct 3 models with 1,2,3 neurons in the hidden layer respectively.

a. Plot each neural net model.

b. Calculate area under ROC and plot the ROC curve for each model (Use test data).

c. Choose the best model with respect to area under ROC curves.

d. Choose threshold by using the ratio of (TP+TN)/(TP+TN+FP+FN) (Accuracy ratio) and report the ratio.

(hint. Use neuralnet & ROSE packages)

PART E Decision Tree

Create a decision tree model. Use same dataset (Logit1.csv) with same test and train data you created for previous question.

a. Plot fitted and actual data for test sample.

b. Plot the decision tree model.

c. Use 0.75 as the threshold, create truth table. Also report accuracy ratio.

d. Calculate Roc and plot the ROC curve.

(hint. Use tree & ROSE package)