

# Stat 479 Presentation: Loan Status Classification

Zongyan Wang

Dept. of Statistics

May 2, 2016

# Overview

- Data overview:  
24 columns, 30,000 rows. We are interested in the last column, which is a categorical variable (1: default, 0: not default)
- X2: Gender, X3: Education, X4: Marital status are categorical variables, transfer them to multiple columns.
- No missing data, the data is very clean
- The sampling is a little bit biased, 78% of data is non-default observation.

# Logistic Regression

- Fit a logistic regression  $\text{Default} \sim$  .
- The in-sample error is 0.1884482759. Nice prediction accuracy!
- Anything missed?

# Logistic Regression

The model is not as good as we thought!

- 1 The error type of Prediction = 0, True = 1 is 0.7578832345

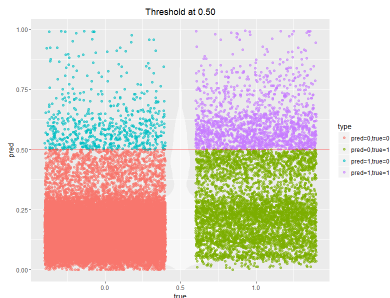


Figure: Prediction Distribution of Logistic Regression

# Neural Network

Neural Network is a very basic approach in Financial area these days, I would like to fit a neural network with a layer(10) to fit the data. Because of the size of our data, even a one-layer neural network takes more than half an hour in R.

# Neural Network

## Neural Network

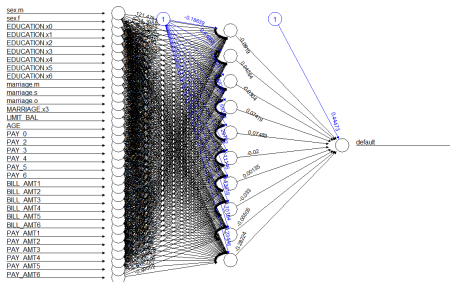


Figure: Neural Network

- The in-sample error is 0.2216206897.

# Neural Network



Figure: Prediction Distribution of Logistic Regression

- The error type of Prediction = 0, True = 1 is 1

# Neural Network

Use ROC to improve the performance.

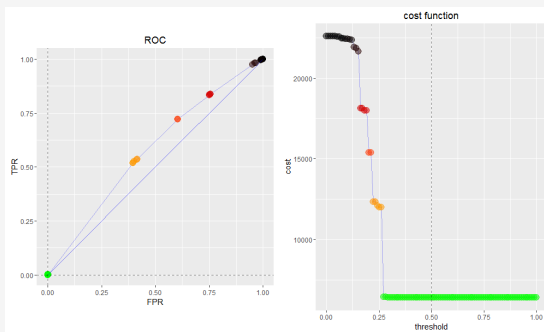


Figure: ROC and Cost function of Neural Network

- The best threshold is every threshold larger than .25 according to the cost function



# Kernel Methods

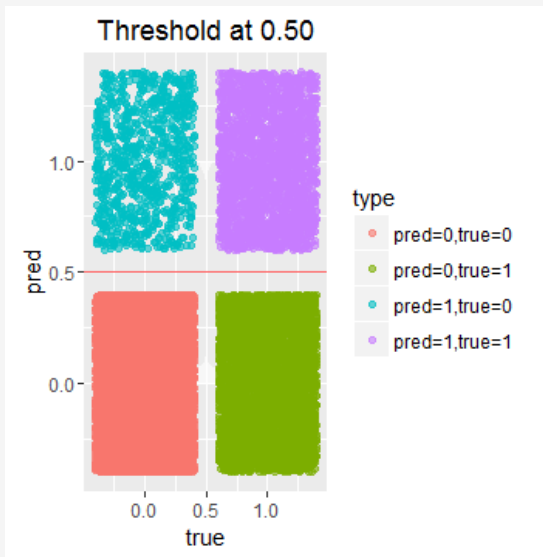
Three kernels have been considered

- Linear kernel gives an in-sample error 0.190276
- Polynomial kernel with degree 6 gives an in-sample error 0.219724
- Gaussian kernel gives an in-sample error 0.174931

The error type of Prediction = 0, True = 1 is smaller, the number of Support Vectors are large

- The error type of Prediction = 0, True = 1 is 0.6544633.
- 12555 observations are selected to be Support Vectors

# Kernel Methods



# Kernel Methods

## Cross Validation and Apply to Test Data

Use Cross-Validation to choose parameter C:

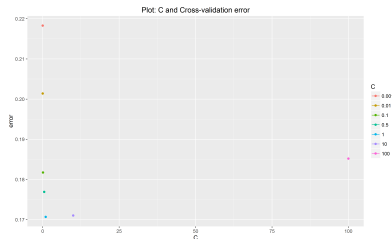


Figure: Cross-validation to select C

- Apply our method to Test set, the error is 0.184.
- The error type of Prediction = 0, True = 1 is 0.6359447

# The End

**Thank you!**