

HW6-Summary

March 2020

1 Credit Card Default Prediction Using TensorFlow (Part-1 Deep Neural Networks)

1.1 Introduction

The authors build a neural network that is capable of predicting the risk of default using the available data using TensorFlow. Because there was no indication of obvious pattern in the data, they start with a neural network with at least 2 hidden layers. The neural network was trained with a random sample of 23000 samples with batch gradient descent. Each batch was 100 samples and for this example we used 9000 iterations.

1.2 Data

The data set used in this example is a real world 30K samples data set of 24 features associated with users and their credit cards default status. The 24 features include demographic info (age, gender, marital status etc.), credit limit, past payment details and other relevant information. The full details of the data set can be found later in this post. For this analysis, they decided not to clean the data in any way. We assumed that the data is clean and doesn't have any problems associated with dirty data sets.

1.3 Results

This neural network can predict the risk of default with an efficiency of almost 0.8. The performance can be further tuned with hyper parameter optimization without overfitting. They see a clear pattern between higher default risk associated with lower limit balances. Also, there seems to be some correlation between the education level and the risk of default. For the age risk, the neural network learnt that the density of the risk is the highest in the 20's and 30's age bracket. The neural network was not able to find any correlation between marriage status and the default risk.

1.4 Link

Results

GitHub

2 Default Risk using Deep Learning

2.1 Introduction

The objective of this project is to use historical loan application data to predict whether or not an applicant will be able to repay a loan. This is a standard supervised classification task. The label is a binary variable, 0 for "will repay loan on time", 1 for "will have difficulty repaying loan". They use deep neural network with 4 layers having 2 hidden layer and each hidden layer having 80 neurons. Dropout layer is added between the 2 hidden layers and between second hidden layer and the output layer to prevent overfitting. The input layer has a dimension of 242, the 2 hidden layers has rectifier as its activation function with 80 neurons each. The output layer has sigmoid activation function as we are doing binary classification. The Deep neural network is compiled with Adam optimizer, Binary-crossentropy loss function and the evaluation metrics is accuracy. They also use deep neural network with 5 layers having 3 hidden layer, where first and second hidden layer has 80 neurons and third hidden layer has 40 neurons. Dropout layer is added between the hidden layers and between third hidden layer and the output layer, Batch normalization layer is inserted between third hidden layer and output layer to prevent overfitting. They use Accuracy, Confusion Matrix, Precision, Recall, True Negative Rate (TNR), False Discovery Rate (FDR), Gain Chart, Lift Chart, K-S Chart, ROC — AUC chart to evaluate the models.

2.2 Data

The data is provided by Home Credit, a service dedicated to provided lines of credit (loans) to the unbanked population. Predicting whether or not a client will repay a loan or have difficulty is a critical business need, and Home Credit wants to unlock the full potential of their data to see what sort of machine learning/deep learning models can be develop to help them in this task. They use Label Encoding for any categorical variables with only 2 categories and One-Hot Encoding for any categorical variables with more than 2 categories. They do statistical analysis and visualization for both numeric and categorical variables.

2.3 Link

Results

Data

Github

3 Predicting Recovery Rates of Defaulted Bonds Using a Neural Network

3.1 Introduction

This is a graduation thesis from UCLA where the authors give a very comprehensive description of models in recovery rate prediction and introduce the application of artificial neural network. This study focuses on the estimation of recovery rates applicable in the framework of the advanced internal ratings-based approach. They evaluate the use of an artificial neural network to predict recovery rates of defaulted corporate bonds. In the neural network, the authors use gradient descent to minimize the error function and sigmoid function as the activation function. The authors fix the number of hidden layer to a single one, and uses 18 neurons in the hidden layer.

3.2 Data

This study uses defaulted bonds from the "Moody's Default Recovery Database". The sample consists of 415 American corporate bonds that defaulted between 1990 and 2016. The distribution of the recovery rates in the database follows a multimodal beta distribution. Lower recovery rates are more frequent and the density decreases as the recovery rates increase. In order to predict the recovery rates, they use three types of explanatory variables: bond-specific, macroeconomic and firm-specific.

3.3 Results

The author uses the Connection Weight method to define the relative importance of an input of a single hidden layer network. Debt seniority, sector and default rates are the most impactful parameters of the model. However, cash flow to total liabilities and the coupon does not have any effect on the recovery rates in the neural network. The technology sector, the default rate and the size of the firm has a negative impact on recovery rates.

3.4 Link

Paper