Deep Learning Charity Funding Predictor Report

Christine Li
`hli94016@usc.edu`

Abstract

1 Overview

Given the background of the non-profit foundation Alphabet Soup wants to create an algorithm to predict whether or not applicants for funding will be successful, we have received a CSV containing more than 34,000 organization that have received funding from Alphabet Soup over the years. With such charity funding information, we would like to compile, train and evaluate a model that can utilize the information inside the CSV file and predict if the applicant of funding will be successful.

2 Exploratory Data Analysis

2.1 Data pre-processing

This dataset contains 34,299 records with 11 variables:

- S_CONSIDERATION
- NAME
- APPLICATION_TYPE
- AFFILIATION
- CLASSIFICATION
- USE_CASE
- ORGANIZATION
- STATUS
- INCOME_AMT
- EIN
- ASK_AMT

With the goal of predicting whether or not applicants for funding will be successful, IS_SUCCESSFUL variable is considered to be the target for my model.

By looking at the 11 variables above, we found that both EIN and NAME variables are string and do not contain any meaningful information to the model, they will not be included into the features for my model. Besides that, other 9 variables will be included as features for my model.

As mentioned above, EIN and NAME variable are neither targets nor features and should be removed from the input data during data pre-processing.

3 Compiling, Training and Evaluating the Model

In this neural network model, we have used relu and sigmoid as activation and 4 layers with different units assigned to the hidden layer of neural network model.

The target model performance is 75%, however, the model we built was not able to achieve this performance.

During the experiment, we use the baseline model that do not have any layer and the epochs number is 50. the accuracy we got for the baseline is 71.45%.

There are a few steps we have try to increase the model performance:

- Increase the epochs number from 50 to 100
- Add one sigmoid layer to the neural network model
- Add another sigmoid layer with 20 units as parameter.
- Add a relu layer with 30 units as parameter
- Use Random Forest model as it is a classifier problem.

In the end, the model we built with the best performance of achieving 72.5% accuracy

4 Summary

During this experiment, we have tried different ways to improve the accuracy of predicting whether or not applicants for funding will be successful. We find that if the candidate with top 5 classification (C1000, C2000, C1200, C3000,

C2100), or the candidate with top 5 Application type (T3, T4, T6, T5, T19), their funding application are more likely to be successful.