Mini-Project 2: Yelp Business Rating Prediction using Tensorflow

Version 1.0

CSC215, Fall 2018

Oct 8th, 2018

Prepared by: Chandini Nagendra

Siddharth Chittora

Contents

[1. Problem Statement 2](#_Toc526176076)

[2. Methodology 3](#_Toc526176077)

[3. Experimental Results and Analysis 3](#_Toc526176078)

[4. Task Division 3](#_Toc526176079)

[4.1. Chandini Nagendra: 3](#_Toc526176080)

[4.2. Siddharth Chittora 3](#_Toc526176081)

[5. Project Reflection 3](#_Toc526176082)

[6. Additional Features 3](#_Toc526176083)

# Problem Statement

In this project, we aim to predict a business's stars rating using the reviews of that business and review count based on neural network implementation in Tensorflow. This project is twofold:

Task 1: Consider this problem as a regression problem. Compare the RMSE of the BEST **Tensorflow regression neural network model** you obtained with that of **regression model** you achieved in the last project.

Task 2: Consider this problem as a classification problem. Compare the accuracy of the BEST **Tensorflow classification neural network model** you obtained with that of **each classification model** you achieved in the last project.

# Methodology

## Regression Problem

### Linear Regression

show predicted and actual for 10 business

show RMSE

### Logistic Regression

show predicted and actual for 10 business

show RMSE

### Simple Regression

show predicted and actual for 10 business

show RMSE

**Activation**

### Regression with early stopping and modelcheckpoint with relu

show predicted and actual for 10 business

show RMSE

### Regression with early stopping and modelcheckpoint with sigmoid

show predicted and actual for 10 business

show RMSE

### Regression with early stopping and modelcheckpoint with tanh

show predicted and actual for 10 business

show RMSE

**Optimizer**

### Stochastic gradient descent optimizer(SGD)

show predicted and actual for 10 business

show RMSE

RMSProp

Adagrad

Adadelta

Adam

Adamax

Nadam

Experiments with Hidden nodes selection in hidden layer

1. trial and error
2. Forward Approach : This approach begins by selecting a small number of hidden neurons. We usually begin with two hidden neurons. After that train and test the neural network. Then increased the number of hidden neurons. Repeat the above procedure until training and testing improved.
3. This approach is opposite of Forward approach. In this approach we start with large number of hidden neurons. Then train and test the NN. After that gradually decrease the number of hidden neurons and again train and test the NN. Repeat the above process until training and testing improved.

3 Rule of thumb method[3] Rule of thumb method is for determining the correct number of neurons to use in the hidden layers, such as the following: • The number of hidden neurons should be in the range between the size of the input layer and the size of the output layer Foram S. Panchal et al, International Journal of Computer Science and Mobile Computing, Vol.3 Issue.11, November- 2014, pg. 455-464 © 2014, IJCSMC All Rights Reserved 462 • The number of hidden neurons should be 2/3 of the input layer size, plus the size of the output layer • The number of hidden neurons should be less than twice the input layer size

## Classification Problem

# Experimental Results and Analysis

# Task Division

## Chandini Nagendra:

## Siddharth Chittora

Discussed together on how to improve the model and came up with the following solution

# Project Reflection

label encoding does not work for tensor flow, when using classification. do one hot coding

with 25 neurons in the 1st layer the r2 score was 60 and with 60 neurons in the 1st and 30 in the second the R2 was 72.

used zscore for tanh

# Additional Features