CSC 215-01 Artificial Intelligence (Fall 2018)

Mini-Project 2: Yelp Business Rating Prediction using Tensorflow

Due at 4:00 pm, Monday, October 8, 2018

Demo Session: class time, Monday, October 8, 2018

1. Problem Formulation

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.

2. Dataset (30 pts)

https://www.yelp.com/dataset/download

This set includes information about local businesses in 10 metropolitan areas across 2 countries. The dataset contain several json files.

3. Additional Requirements

- You are required to split data to training and test. Use training data to train your models and evaluate the model quality using test data. You do not need to do k-fold cross validation.
- Use Sklearn, Keras and Tensorflow to finish this project. Any other Python libraries are also welcome to use.
- Use TF-IDF to do feature extraction from review contents for your models

- If you experience low memory error on your machine when you use ***tfidfVectorzier***, set parameters ***max_df, min_df,*** and ***max_features*** appropriately.
- Do feature normalization.
- You must use EarlyStopping and ModelCheckpoint when training neural networks using Tensorflow.
- Tuning the following hyperparameters when training neural networks using Tensorflow to see how they affect performance
    - **Activation:** relu, sigmoid, tanh
    - **Layers and Neuron Counts**
    - **Optimizer:** adam, sgd, rmsprop, and others

4. Grading breakdown

You may feel this project is described with some certain degree of vagueness, which is left on purpose. In other words, **creativity is strongly encouraged**. Your grade for this project will be based on the soundness of your design, the novelty of your work, and the effort you put into the project.

Use the evaluation form on Canvas as a checklist to make sure your work meet all the requirements.

Implementation	**70 pts**
Your report	**15 pts**
In-class defense	**10 pts**
Additional features (novelty)	**5 pts**

5. Teaming:

Students must work in teams of 2 people. Think clearly about who will do what on the project. Normally people in the same group will receive the same grade. However, the instructor reserve the right to assign different grades to team members depending on their contributions. So you should choose partner carefully!

6. Deliverables:

(1) All your source code in Python Jupyter notebook.

(2) Your report in PDF format, with your name, your id, course title, assignment id, and due date on the first page. As for length, I would expect a report with more than one page. In the report, include two sections (1) "**Task Division and Project Reflection**" and (2) **Additional Features.**

In the section "**Task Division and Project Reflection**", describe the following:

- who is responsible for which part,
- challenges your group encountered and how you solved them
- and what you have learned from the project as a team.

10 pts will be deducted for missing this section.

In the section "**Additional Features**", describe any additional features (if there is any).

All the files must be submitted **by team leader** on Canvas before

4:00 pm, Monday, October 8, 2018

NO late submissions will be accepted.

7. In-class defense:

Each team member must defend your work during the scheduled defense session. Each team have **five minutes** to defend your work in class. In the defense, briefly describe **the basic steps** you took to finish this project by referring audience to the data/code/figures in your Jupyter notebook.

If you implement **additional features (novelty)**, please do mention them to receive credit for novelty.

Failure to show up in defense session will result in **zero** point for the project.