Spring 2022: CSEE5590/490 – Special Topics

Python and Deep Learning Module-2 - ICP-11

Lesson Overview:
In this lesson, we are going to discuss Word Embedding.

Use Case Description:
1. Sentiment Analysis on the IMDB dataset

Source Code:
Provided in your assignment folder and assignment repo. Dataset uploaded to Box [Link]

In class programming:

1. In the code provided (PPT) there are multiple mistakes which stop the code from running successfully.
    a. find those mistakes and explain why they need to be corrected to be able to get the code run.
2. Drop the "unsup" label from the dataset file.
3. Filter the reviews text by:
    a. Removing punctuation characters.
    b. Lower case words.
    c. Reduce the words to their root.
4. Add embedding layer to the model, did you experience any improvement?
5. Does your model suffer from overfitting or underfitting?
    a. If your model is overfitting, how would you prevent this?
6. Apply the code on 20_newsgroup data set we worked in the previous classes.

```
from sklearn.datasets import fetch_20newsgroups
newsgroups_train = fetch_20newsgroups(subset='train', shuffle=True,
categories=categories,)
```

7. Predict over one sample of data and check what will be the prediction for that.

(Extra Credit) (5 points):
1- Add to your model Convolution and Maxpooling layer.

** Follow the IPC rubric guidelines.

Submission Guidelines:

1. Once finished document your code and make sure all parts of the assignments are completed.
2. Push your code to your GitHub repo and update the ReadMe file, add your info, and partner info.
3. Submit the assignment on Canvas.
4. Present your work to TA during class time to prove the execution and complete submission.

After class submission:

1. Once finished document your code and make sure all parts of the assignments are completed.
2. Push your code to your GitHub repo and update the ReadMe file, add your info, and partner info.
3. Submit the assignment on Canvas before the deadline.
4. Record a short video (3~7) minute, proof of execution and complete assignment.
5. Add video link to ReadMe file.

Note: *Cheating, plagiarism, disruptive behavior, and other forms of unacceptable conduct are subject to strong sanctions in accordance with university policy. See detailed description of university policy at the following URL: https://catalog.umkc.edu/special-notices/academic-honesty/*