ASSIGNMENT 3

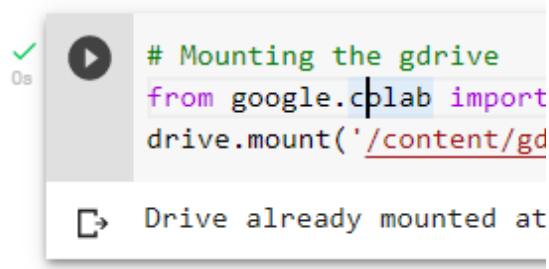
- Manvi Goel(2019472) and Jahnvi Kumari(2019469)

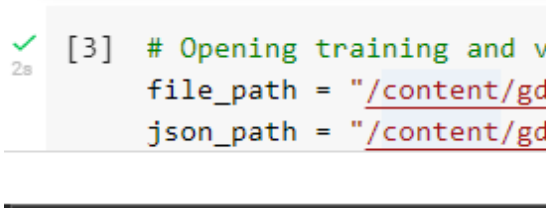
Execution

- Mount the drive.

```
# Mounting the gdrive
from google.colab import drive
drive.mount('/content/gdrive')
```

- Open the training and the validation files.
- Run the cells in the set order, or run all the cells





Model

Methodology

HMM with Viterbi

- Made a dictionary called train_word_tag of the form {Label: {word:count}}.
- Made a bigram tag probability for each tag.
- Make a transition probability table for each tag.
- Trained the model using Viterbi and backtracking.

MLP with Word2Vec

- Make word2vec feature vectors for all words.
- Train using feature vectors and labels
- Evaluate the model using MLP

MLP with Glove

- Make Glove feature vectors for all words.
- for words not present, split words using ("-", "`") and multiply each vector.
- if not possible, make random vectors.
- Train using feature vectors and labels
- Evaluate the model using MLP

Preprocessing

- All the words are turned to lowercase.
- Removing space from sentence ends.
- Label encoding and one hot encoding.
- Inserted <start> tag in Viterbi

Parameters

- The current function uses the test set to calculate the accuracy. Upload the test set, and update the file path to check the accuracy of the test set.
- Parameters of MLP can be changed

Assumptions

- The test file is in txt.
- The words present in the test file but not in the training set have been assigned a very low probability of being present at the start. It has been tagged as NOUN.

Thank You