CS564 - Foundation of Machine Learning
(Read all the instructions carefully and adhere to them.)

Assignment - 4: HMM based POS Tagger

Deadline: 07/11/2019 **Date: 31/10/2019**

Design and implement a Hidden Markov Model (HMM) based Part-of-Speech (POS) tagger implementing Viterbi algorithm. PoS tagging refers to assigning most appropriate sequence of part-of-speech tags to each word in a sentence.

Specifications:
i. Given the Brown PoS tagged corpus for experiment**.**
ii. Dataset format: Each line represents one sentence; Sentences are already tokenized; Words in a line have the format word_tag.
iii. Implement a second-order HMM.
iv. Implement Viterbi algorithm for finding the best state sequence.
v. Implement Forward and Backward recursions for the evaluation problem.
vi. Use Maximum Likelihood Estimation for computing Emission and Transition probabilities.
vii. Make judicious assumption for initialization.

Evaluation:
Perform 3-fold cross validation.
ii. Calculate the overall accuracy, precision, recall and F-score.
iii. Calculate the tag-wise accuracy, precision, recall and F1-score.
iv. Draw the confusion matrix (Each element A_{ij} of matrix A denotes the number of times tag i classified as tag j).

Submission guidelines:
1.Markings will be based on the correctness and soundness of the outputs. Marks will be deducted in case of plagiarism.
2. Proper indentation and appropriate comments (if necessary) are mandatory.
3. You should zip all the required files and name the zip file as roll_no_of_all_group_members.zip, eg. 1601cs11_1601cs03_1621cs05.zip.
4. Upload your assignment (the zip file) in the following link:
https://www.dropbox.com/request/BGhibCkCcr7SQdfiS4tc

For any queries regarding this assignment contact:
Chanchal Suman(chanchaliitp@gmail.com)