Assignment-Discussion Vector Based POS Tagging

<name-1>, <roll no.-1>
<name-2>, <roll no.-2>
…
<date>

Problem Statement: **part 1**

- Given a sequence of words, produce the POS tag sequence

- Technique to be used: HMM-Viterbi-vector (vector based; the whole corpus is corpus of word vectors which replace words)

- Use Universal Tag Set (12 in number); <list the tags>

- 5-fold cross validation

- Compare with HMM-Viterbi-symbolic

Problem Statement: part 2

- Given a sequence of words, produce the POS tag sequence

- Technique to be used: word2vec vectors, FFNN and BP (a slide on FFNN-BP architecture is a must)

- Use Universal Tag Set (12 in number); <list the tags>

- 5-fold cross validation

- Compare with HMM-Viterbi-symbolic

Overall performance

- Precision
- Recall
- F-score (3 values)
  - F1-score
  - F0.5-score
  - F2-score
- For both part 1 and 2 and compare
- Also Compare with HMM-Viterbi-symbolic

Per POS performance

- Tag1: P, R, F1
- Tag2: P, R, F1
- …
- …
- Tag-12: P, R, F1

- Compare all three models

Confusion Matrix (12 X 12) (can give heat map) (compare all 3 models)

Interpretation of confusion (error analysis)

- <list maximal confusions; which tag is confused with which tag most>

- <try giving reasons>

- Compare all 3 models

Data Processing and Data Sparsity

- <Describe how you obtained the word vectors>

- For solving the problem of unseen words use cosine similarity of vectors