

## Project 4 - POS Tagging

---

`./viterbi.py input_hmm test_file output_file`

I'm providing `hmm.txt`, which is in the following format (italics are replaced with values in the file):

`state_num=number of states`

`sym_num=number of POS tags`

`\init`

`tag  $P(\text{tag} \mid \text{BOS}) \log_{10}(P(\text{tag} \mid \text{BOS}))$`

`...`

`\transition`

`tag1 tag2  $P(\text{tag2} \mid \text{tag1}) \log_{10}(P(\text{tag2} \mid \text{tag1}))$`

`...`

`\emission`

`tag word  $P(\text{word} \mid \text{tag}) \log_{10}(P(\text{word} \mid \text{tag}))$`

`...`

- \* If there is no transition probability from tag1 to tag2, then it is impossible to go from tag1 to tag2.
- \* If there is no transition probability from tag to word, then it is impossible for that tag to generate that word.

Each line of the `test_file` contains a single sentence.

The format of the `output_file` should be, one sentence per line:

`word1/tag1 word2/tag2 word3/tag3...`