

Part-of-speech tagging

Use the first 10k tagged sentences from the Brown corpus to **generate the components of a part-of-speech hidden markov model**: the transition matrix, observation matrix, and initial state distribution. Use the universal tagset:

```
nltk.corpus.brown.tagged_sents(tagset='universal')[:10000]
```

Also hang on to the mappings between states/observations and indices. Include an OOV/UNK observation and smoothing everywhere.

Using the provided Viterbi implementation, **infer the sequence of states for sentences 10150-10152 of the Brown corpus**:

```
nltk.corpus.brown.tagged_sents(tagset='universal')[10150:10153]
```

and **compare against the truth. Explain why your POS tagger does or does not produce the correct tags.**

You may work in a group of 1 or 2. Submissions will be graded without regard for the group size. You should turn in a document (`.txt`, `.md`, or `.pdf`) answering all of the **red** items above. You should also turn in Python scripts (`.py`) for *each* of the **blue** items. Unless otherwise specified, you may use only `numpy` and the **standard library**.