Q1

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 observation and smoothing everywhere.

Q2

Implement a function `viterbi()` that takes arguments:

1. `obs` - the observations [list of ints]
2. `pi` - the initial state probabilities [list of floats]
3. `A` - the state transition probability matrix [2D numpy array]
4. `B` - the observation probability matrix [2D numpy array]

and returns:

1. `states` - the inferred state sequence [list of ints]

Do everything in log space to avoid underflow.

Q3

Infer the sequence of states for senteces 10150-10152 of the Brown corpus:

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

and compare against the truth.

You should work independently. You may use only built-in Python modules and numpy. Submit your solutions as a Jupyter notebook (.ipynb file) along with any auxiliary Python files, in a .zip archive.