Tutorial 4: POS tagging and HMM

Q1. Find tagging errors in each of the following sentences that are tagged with the Penn Treebank tagset. You may get help from online demos of POS tagging services.

1. How/WRB do/MD I/PRP get/VB to/TO Singapore/NN
2. Do/VBP you/PRP have/VB any/DT vacancies/NN
3. This/DT room/NN is/VBZ too/JJ noisy/JJ
4. Can/VB you/PRP give/VB me/PRP another/DT room/NN

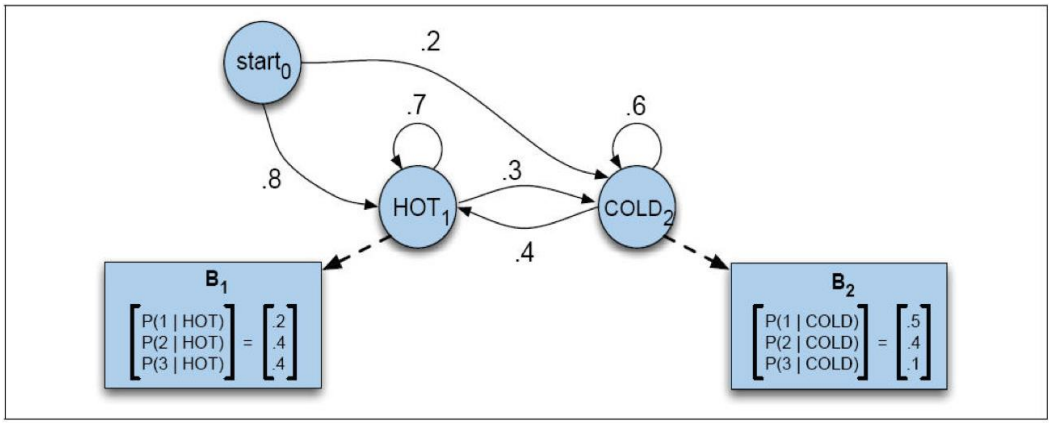
Q2. Compute the best tag sequence for "I want to race" using the Viterbi algorithm with the provided HMM parameters, i.e., the transition probability and the word likelihood probabilities.

	VB	TO	NN	PPSS
<s>	.019	.0043	.041	.067
VB	.0038	.035	.047	.0070
TO	.83	0	.00047	0
NN	.0040	.016	.087	.0045
PPSS	.23	.00079	.0012	.00014

	I	want	to	race
VB	0	.0093	0	.00012
TO	0	0	.99	0
NN	0	.000054	0	.00057
PPSS	.37	0	0	0

Q3. Run the Viterbi algorithm with the HMM in the figure below to compute the most likely weather sequences for each of the two following observation sequences. Note: You may consider using HMM packages for computation, e.g., https://pypi.org/project/hmmlearn/

- Sequence: 312312312
- Sequence: 311233112



Q4. The task of **negation scope detection** is to extract the parts of a sentence that is being negated. For example, in the sentence "I have not submitted my assignment", the negation scope is "submitted my assignment".

Formulate this problem as a sequence labelling task, and discuss how to apply Hidden Markov Model (HMM) to solve this problem. Clearly state the probabilities that need to be learned by the HMM.