Part of Speech Tagging with Hidden Markov Models

1 Sequence Probability

You are trying to compute the log probability of the tag sequence “`V P V`” given the sequence of words “`ski on snow`”. This would be represented as

$$\log(P(\texttt{V P V}|\texttt{ski on snow}))$$

Remember that with the assumptions we will make in our HMM tagger,

$$\log\left(P(t_1...t_n|w_1...w_n)\right) = \log\left(\prod_{i=1}^{n} P(w_i|t_i)P(t_i|t_{i-1})\right) = \sum_{i=1}^{n}\left(\log P(w_i|t_i) + \log P(t_i|t_{i-1})\right)$$

We will state that $P(t_1|t_0) = P(t_1| < s >)$, where `<s>` is a special tag indicating the start of the sequence.

1.1 Which probabilities?

Without writing out any numerical probabilities, complete equation below to show the conditional probabilities you must add.

$\log\left(P(\texttt{ski}|\texttt{V})\right) + \log\left(P(\texttt{V}|\texttt{<s>})\right) +$

1.2 Plugging in numbers

Imagine your HMM is defined by the following initial, transition, and emission **log** probabilities. The initial probabilities represent $P(t| < \mathrm{s} >)$.

tag	V	N	P
log prob	-3	-3	-3

Table 1: Initial **log** probabilities

	second tag		
first tag	V	N	P
V	-4	-2	-2
N	-3	-2	-1
P	-5	-2	-4

Table 2: Transition **log** probabilities

	word		
tag	ski	on	snow
V	-6	$-\infty$	-5
N	-5	$-\infty$	-3
P	$-\infty$	-1	$-\infty$

Table 3: Emission **log** probabilities

Use the probabilities above to compute

$$\log(P(\texttt{V P V}|\texttt{ski on snow}))$$

2 Viterbi Algorithm

2.1 Completing the *viterbi* table

Using the initial, transition, and emission **log probabilities** from the previous page, fill out the table representing the execution of the Viterbi algorithm. Your final answers in each box should be in the form of log probabilities (for *argmax*, log probabilities are suitable). Keep track of backpointers, either by writing arrows or writing the tag that led to the maximum probability in each box you fill out.

	ski	on	snow
V			
N			
P			

2.2 Finding the final sequence

What is the most probable POS sequence identified by the Viterbi algorithm?