HW7: Part of Speech Tagging
Solution to written problems

The following tagging model has the two words *'boxes'* and *'books'*, and the two POS tags *'noun'* and *'verb'* plus the end-of-sentence delimiters $\triangleright$ and $\triangleleft$.

$p(\text{NOUN}\mid\triangleright) = 1/2$	$p(\text{VERB}\mid\triangleright) = 1/2$	$p(boxes\mid\text{NOUN}) = 1/2$
$p(\text{NOUN}\mid\text{NOUN}) = 1/2$	$p(\text{VERB}\mid\text{NOUN}) = 1/6$	$p(boxes\mid\text{VERB}) = 3/4$
$p(\text{NOUN}\mid\text{VERB}) = 1/2$	$p(\text{VERB}\mid\text{VERB}) = 1/6$	$p(books\mid\text{NOUN}) = 1/2$
$p(\triangleleft\mid\text{NOUN}) = 1/3$	$p(\triangleleft\mid\text{VERB}) = 1/3$	$p(books\mid\text{VERB}) = 1/4$

1 Total probability of "boxes books"

To calculate the total probability of an output sequence under an HMM we need to ***marginalize*** over the hidden state sequences. In other words, we calculate each possible tag sequences that could generate the output sequence in question and *sum up* their individual probabilities.

In this case there are four possible tag sequences: VV, VN, NV, and NN.

$$\begin{aligned} P(\text{boxes books}) = {} & P(<\text{boxes books}>, <\triangleright, V, N, \triangleleft>) + \\ & P(<\text{boxes books}>, <\triangleright, V, V, \triangleleft>) + \\ & P(<\text{boxes books}>, <\triangleright, N, V, \triangleleft>) + \\ & P(<\text{boxes books}>, <\triangleright, N, N, \triangleleft>) \end{aligned}$$

We can calculate them individually by taking the product of the respective transition (σ) and emission (τ) probabilities

$$\begin{aligned} P(<\text{boxes books}>, <\triangleright, V, N, \triangleleft>) &= \sigma_{\triangleright,V} \ * \ \tau_{V,boxes} \ * \ \sigma_{V,N} \ * \ +\tau_{N,books} \ * \ \sigma_{N,\triangleleft} \\ &= \frac{1}{2} * \frac{3}{4} * \frac{1}{2} * \frac{1}{2} * \frac{1}{3} \\ &= \frac{1}{32} \\ &= 0.03125 \end{aligned}$$

$$\begin{aligned} P(<\text{boxes books}>, <\triangleright, V, V, \triangleleft>) &= \sigma_{\triangleright,V} \ * \ \tau_{V,boxes} \ * \ \sigma_{V,V} \ * \ +\tau_{V,books} \ * \ \sigma_{V,\triangleleft} \\ &= \frac{1}{2} * \frac{3}{4} * \frac{1}{6} * \frac{1}{4} * \frac{1}{3} \\ &= \frac{1}{192} \\ &\approx 0.00521 \end{aligned}$$

$$
\begin{aligned}
P(<\text{boxes books}>, <\triangleright, N, V, \triangleleft>) &= \sigma_{\triangleright,N} \;*\; \tau_{N,boxes} \;*\; \sigma_{N,V} \;*\; +\tau_{V,books} \;*\; \sigma_{V,\triangleleft} \\
&= \frac{1}{2} * \frac{1}{2} * \frac{1}{6} * \frac{1}{4} * \frac{1}{3} \\
&= \frac{1}{288} \\
&\approx 0.00347
\end{aligned}
$$

$$
\begin{aligned}
P(<\text{boxes books}>, <\triangleright, N, N, \triangleleft>) &= \sigma_{\triangleright,N} \;*\; \tau_{N,boxes} \;*\; \sigma_{N,N} \;*\; +\tau_{N,books} \;*\; \sigma_{N,\triangleleft} \\
&= \frac{1}{2} * \frac{1}{2} * \frac{1}{2} * \frac{1}{2} * \frac{1}{3} \\
&= \frac{1}{48} \\
&\approx \; 0.02083
\end{aligned}
$$

Thus plugging in these values for each tag sequence to the original formula we get:

$$
\begin{aligned}
P(\text{boxes books}) &= P(<\text{boxes books}>, <\triangleright, V, N, \triangleleft>) \; + \; P(<\text{boxes books}>, <\triangleright, V, V, \triangleleft>) \; + \\
&\quad P(<\text{boxes books}>, <\triangleright, N, V, \triangleleft>) \; + \; P(<\text{boxes books}>, <\triangleright, N, N, \triangleleft>) \\
&= \frac{1}{32} + \frac{1}{192} + \frac{1}{288} + \frac{1}{48} \\
&= \frac{35}{576} \\
&\approx 0.06076
\end{aligned}
$$

This means that if we used the HMM as a generative language model, the likelihood of any particular run generating the exact output sequence *"boxes books"* is ≈ 0.061

2 Probability of the most likely tag sequence

We've already calculated essentially everything we need to answer this in the first part. There are four possible tag sequences with the associated probabilities:

$$
\begin{aligned}
P(<\text{boxes books}>, <\triangleright, V, N, \triangleleft>) &= 0.03125 \\
P(<\text{boxes books}>, <\triangleright, V, V, \triangleleft>) &\approx 0.00521 \\
P(<\text{boxes books}>, <\triangleright, N, V, \triangleleft>) &\approx 0.00347 \\
P(<\text{boxes books}>, <\triangleright, N, N, \triangleleft>) &\approx 0.02083
\end{aligned}
$$

The most probable tag sequence is thus $<\triangleright, V, N, \triangleleft>$ with probability 0.03125

3 Zeroing out

Q: Now suppose for some HMM (not the one described above) that when applied to the sequence $\mathbf{x}$, $P(\mathbf{x}|\mathbf{y}) = 0$ for all $\mathbf{y} =< ..., y_i = a, ... >$. That is, any sequence of states that goes through state a at position i assigns zero probability to the string $\mathbf{x}$. Does it follow that $\tau_{a,x_i} = 0$? Why or why not?

A: While it is the case that *if* $\tau_{a,x_i} = 0$ then every sequence passing through state a at position i will have probability zero, however this is not the only thing that would cause that zero probabilities for all $\mathbf{y} =< ..., y_i = a, ... >$.

For instance, it is also conceivable that something *else* is zeroing all of those probabilities: e.g. perhaps $\sigma_{a,\cdot}$ is zero.

4 Unigram tagger

For any given sentence, the Viterbi sequence of labels may not be the sequence of labels that *individually* have the highest label probabilities (if considering words independently of one another). This is because, the most likely sequence of labels under our HMM (the Viterbi sequence) is calculated based on both the highest individual label probabilities (i.e. the unigram POS tags) ***as well as*** the transitions between states.

5 Majority Viterbi

This in fact is a "coincidence." Which is to say that it's not a necessary property of HMMs. While many output sequences are in practice highly ambiguous, and an HMM may distribute probability over a wide number of possible state sequences, not *every* output sequence is highly ambiguous.

To take a concrete example, notice that the first HMM in the assignment with the "boxes books" sequence demonstrates this. The most likely sequence of labels (V,N) does have a majority of the total sentence probability.