This list contains exercises of the type you will find in an exam for the course *Natuurlijke Taalmodellen en Interfaces*.

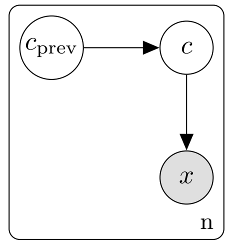
Contents

Points

Question:	1	2	3	4	5	6	7	Total
Points:	2	1	2	2	2	1	3	13

1 Hidden Markov models

The hidden Markov model (HMM) extends the Markov model with word categories. The graphical model below specifies the conditional independence assumptions of the HMM, where



X is a random word from a vocabulary of v words and C is a random word category (or tag) from a vocabulary of t tags. There are two types of cpds in the HMM. *Transition distributions* used to generate a tag given the tag of the previous word:

$$C|C_{\text{prev}} = c_{\text{prev}} \sim \text{Cat}(\lambda_1^{(c_{\text{prev}})}, \ldots \lambda_t^{(c_{\text{prev}})})$$

And *emission distributions* used to generate a word given its tag:

$$X|C = c \sim \text{Cat}(\theta_1^{(c)}, \ldots, \theta_v^{(c)})$$

The joint probability for a sentence x_1^n and tag-sequence c_1^n given length $N = n$ factorises

$$\begin{aligned} P_{X_1^n C_1^n|N}(x_1^n, c_1^n|n) &= P_{X_1^n C_1^n|N}(x_1^n, c_1^n|n) \\ &= \prod_{i=1}^{n} P_{C|C_{\text{prev}}}(c_i|c_{i-1}) P_{X|C}(x_i|c_i) \end{aligned}$$

in terms of transition and emission probabilities. Assessing the probability of a sentence, regardless of tag sequence, requires marginalisation

$$P_{X_1^n|N}(x_1^n|n) = \sum_{c_1=1}^{t} \cdots \sum_{c_n=1}^{t} P_{X_1^n C_1^n|N}(x_1^n, c_1^n|n)$$

1. (2 points) Express the joint probability for a sentence x_1^n and tag-sequence c_1^n given their length as a function of transition and emission parameters.

2. (1 point) The HMM has ____________ transition distributions, each of which has ____________ parameters, it also contains ____________ emission distributions, each of which has ____________ parameters, and therefore, the total representation cost of the HMM (in big-O-notation) is ____________.

3. (2 points) Write the generative story of the HMM (you may assume appropriate padding exists).

4. (2 points) Consider the following transition and emission distributions.

	$C=1$	$C=2$	$C=3$
$C_{\text{prev}}=0$	$\lambda_1^{(0)}$	$\lambda_2^{(0)}$	$\lambda_3^{(0)}$
$C_{\text{prev}}=1$	$\lambda_1^{(1)}$	$\lambda_2^{(1)}$	$\lambda_3^{(1)}$
$C_{\text{prev}}=2$	$\lambda_1^{(2)}$	$\lambda_2^{(2)}$	$\lambda_3^{(2)}$
$C_{\text{prev}}=3$	$\lambda_1^{(3)}$	$\lambda_2^{(3)}$	$\lambda_3^{(3)}$

	$X=1$	$X=2$	$X=3$	$\dots$	$X=v$
$C=1$	$\theta_1^{(1)}$	$\theta_2^{(1)}$	$\theta_3^{(1)}$	$\dots$	$\theta_v^{(1)}$
$C=2$	$\theta_1^{(2)}$	$\theta_2^{(2)}$	$\theta_3^{(2)}$	$\dots$	$\theta_v^{(2)}$
$C=3$	$\theta_1^{(3)}$	$\theta_2^{(3)}$	$\theta_3^{(3)}$	$\dots$	$\theta_v^{(3)}$

Transition distributions (left) and emission distributions (right)

We can use an HMM model defined with these cpds to find the best possible way to tag an input sentence $\langle x_1, x_2, x_3 \rangle$. The table below shows 3 cells used to compute the Viterbi recursion $\alpha(i,j)$ where i corresponds to the position in the sentence and j corresponds to a tag in the POS tagset. What is the value of the Viterbi entry $\alpha(i=3, j=1)$?

	$i=1$	$i=2$	$i=3$
$C=1$	$\lambda_1^{(0)}\theta_{x_1}^{(1)}$		?
$C=2$	$\lambda_2^{(0)}\theta_{x_1}^{(2)}$		
$C=3$	$\lambda_3^{(0)}\theta_{x_1}^{(3)}$		

Viterbi table $\alpha(i,j)$: assume $j=0$ to correspond to the BoS tag.

5. (2 points) Complete the forward recursion below, where θ denotes emission parameters and λ denotes transition parameters, i ranges over input positions, j and p range over tags in the the tagset.

$$\alpha(i,j) = \begin{cases} \theta^j_{x_i} \times \lambda^{(0)}_j & \text{if } i = 1 \\ \boxed{} \times \sum_{p \in \{1,\ldots,t\}} \boxed{} & \text{otherwise} \end{cases}$$

6. (1 point) Assuming n is the length of a sentence and $t+1$ represents the EoS tag, which cell in the forward table contains the marginal probability of the sentence?

7. Consider the tagged sequences below where the **first sequence occurs n_1 times**, the **second sequence occurs n_2 times**, and the **third sequence occurs n_3 times**.

(1)
$$\begin{tikzcd}
\text{BoS} \arrow[r, dash] & \text{A} \arrow[r, dash] \arrow[d, dash] & \text{EoS} \arrow[d, dash] \\
 & \text{x} & \text{EoS}
\end{tikzcd}$$

(2)
$$\begin{tikzcd}
\text{BoS} \arrow[r, dash] & \text{A} \arrow[r, dash] \arrow[d, dash] & \text{B} \arrow[r, dash] \arrow[d, dash] & \text{EoS} \arrow[d, dash] \\
 & \text{x} & \text{y} & \text{EoS}
\end{tikzcd}$$

(3)
$$\begin{tikzcd}
\text{BoS} \arrow[r, dash] & \text{A} \arrow[r, dash] \arrow[d, dash] & \text{B} \arrow[r, dash] \arrow[d, dash] & \text{C} \arrow[r, dash] \arrow[d, dash] & \text{A} \arrow[r, dash] \arrow[d, dash] & \text{EoS} \arrow[d, dash] \\
 & \text{x} & \text{y} & \text{t} & \text{z} & \text{EoS}
\end{tikzcd}$$

(a) (1 point) Estimate by maximum likelihood the transition distribution given that the previous category is 'A'.

(b) (1 point) Estimate by maximum likelihood the emission distribution given that the category is 'A'.

(c) (1 point) What is the probability of the second sequence pair, given its length, as a function of maximum likelihood estimates?

Total for Question 7: 3

Assessment

Question	Points	Score
1	2	
2	1	
3	2	
4	2	
5	2	
6	1	
7	3	
Total:	13	