Hidden Markov Models

1 Questions

1. What is common in the concepts of GMMs and HMMs? What is the difference between the concepts of GMMs and HMMs?

2. Describe the basic functionality of a HMM using the diagram on slide 14. Which quantities determine the expected output length of a HMM?

3. Which two matrices determine the characteristics of the HMM in a unique way? How is the HMM model $\boldsymbol{\lambda}$ thus defined?
   In literature, often additionally an initial distribution $\boldsymbol{\pi}$ is given, which defines for each model state the probability that the HMM starts in this state. How does the vector $\boldsymbol{\pi}$ look like for a HMM definition in the style of this lecture?

4. Describe the three problems / tasks of a HMM in your own words. Name the applications of these problems.

5. What is a forward probability? Describe the forward algorithm in a few words. Explain shortly the three main parts (initialization, recursion, termination) of the algortihm. Could the algorithm also be applied backwards?

6. Describe the Viterbi algorithm in your own words. You can refer to slide 32. What meaning does the variable $t_i(n)$ have?

2 Answers

1. Both HMMs and GMMs model emission probabilities, but only the HMM also models a temporal behaviour

2. On slide 14 is a HMM depicted, that has four states (S_0 , S_1 , S_2 , S_3) with initial state S_0 and final state S_3. The transition probabilities are given with $a_{i,j}$. The emission probabilities are expressed by $b_i(\boldsymbol{x})$. If the transition probabilities to the final state (here: a_{13} and a_{23}) are close to 1, the expected output length is small.

3. The defining matrices are the transition matrix $\boldsymbol{A}$ and the emission matrix $\boldsymbol{B}$. The HMM can thus be defined by $\boldsymbol{\lambda} = (\boldsymbol{A}, \boldsymbol{B})$.
   Because in the context of this lecture, the HMM always starts in the dedicated initial state S_0, the initial distribution vector can be written as $\boldsymbol{\pi} = (1, 0, \ldots, 0)$.

4. Evaluation, decoding, and estimation (slide 23). An examples for evalution is speech recognition. Using the evaluation, the probability that an HMM generates a given feature vector sequence (of an utterance) can be determined, i.e., how well the HMM matches the utterance. An example for estimation is model training, where the HMM parameters are estimated on the basis of a database of feature vector sequences.

5. The forward probability $f_i(n)$ (is defined as the probability that after n time steps the HMM is in the state S_i and generated the feature sequence $\boldsymbol{X}^{(n)}$ on its way. The three steps are explained on slide 29. Yes, in order to solve the estimation problem, a backward algorithm will be introduced.

6. Initialization, recursion, and termination. In the variable $t_i(n)$, the information to backtrack the optimal path are stored. An illustration can be found on slide 44.