ML E2022 - Week 11 - Theoretical Exercises

Hidden Markov Models

Exercise 1: Questions to slides *Hidden Markov Models - Training*:

1. Consider the simple "weather-HMM" with a transition diagram as shown on slide 3. Assume that we do not know the model parameters i.e. the start-, transition-, and emission-probabilities, but that we are given two pairs of $(\mathbf{X}, \mathbf{Z})$ as training data.

   These pairs are: (`HHLLLHHHHLLLLHH` , `SSSRRSSSRRRRSSS`) and (`LLHHLLLHHHHLLHH` , `RRRSSRRRSSSRRRS`), where H and L are the two states of the model, and S and R are the two emissions sunshine and rain.

   Use Training-by-Counting to set the model parameters according to this training data.

   Solution:

   We count the number of starts, emissions, and transitions in each of two pairs of $(\mathbf{X}, \mathbf{Z})$, and set the probabilities accordingly:

   `HHLLLHHHHLLLLHH`

   `SSSRRSSSRRRRSSS`

   and

   `LLHHLLLHHHHLLHH`

   `RRRSSRRRSSSRRRS`

   Counting starts and setting start probabilities:

   #(start in H) = 1

   #(start in L) = 1

   P(start in H) = 1 / (1+1) = 0,5

   P(start in L) = 1 / (1+1) = 0,5

   Counting transitions and setting transition probabilities:

   #(H -> H) = 5 + 5 = 10

#(H -> L) = 2 + 2 = 4

P(H -> H) = 10 / (10+4) = 0,71

P(H -> L) = 4 / (10+4) = 0,29

#(L -> H) = 2 + 3 = 5

#(L -> L) = 5 + 4 = 9

P(L -> H) = 5 / (5+9) = 0,36

P(L -> L) = 9 / (5+9) = 0,64

Counting emissions and setting emission probabilities:

#(S from H) = 7 + 5 = 12

#(R from H) = 1 + 3 = 4

P(S from H) = 12 / (12+4) = 0,75

P(R from H) = 4 / (12+4) = 0,25

#(S from L) = 2 + 1 = 3

#(R from L) = 5 + 6 = 11

P(S from L) = 3 / (3+11) = 0,21

P(R from L) = 11 / (3+11) = 0,79 `

1. Consider Viterbi training as explained on slides 18-19. If a parameter in the initial model Θ^0 is set to zero, i.e. if a particular transition or emission probability is set to zero, then it will remain zero during all the iterations of Viterbi training (if we do not perform pseudo counts). Why?

   Solution:

   If a particular transition or emission probability is set to zero in Θ^0, then this transition or emission cannot be part of the most likely sequence of hidden states determined by Viterbi decoding. This means that the transition or emission will not be seen in the training data we use for training-by-counting to obtain Θ^1, and therefore remains zero in this model, and similarly remains zero in all future models Θ^i that will be generated in the iterative Viterbi-training proces.

1. Explain why you can stop Viterbi training if the Viterbi decoding does not change between two iterations?

   Solution:

Let $\mathbf{Z}_{\text{Vit}}^{i}$ be the most likely explanation of $\mathbf{X}$ under the model $\mathbf{\Theta^{i}}$ as obtained by Viterbi decoding. We use $(\mathbf{X}, \mathbf{Z}_{\text{Vit}}^{i})$ and training-by-counting to obtain $\mathbf{\Theta^{i+1}}$. Let $\mathbf{Z}_{\text{Vit}}^{i+1}$ be the most likely explanation of $\mathbf{X}$ under the new model $\mathbf{\Theta^{i+1}}$. If $\mathbf{Z}_{\text{Vit}}^{i+1} = \mathbf{Z}_{\text{Vit}}^{i}$, then it is clear that next model, $\mathbf{\Theta^{i+2}}$, obtained by training-by-counting on $(\mathbf{X}, \mathbf{Z}_{\text{Vit}}^{i+1}) = (\mathbf{X}, \mathbf{Z}_{\text{Vit}}^{i})$ will be the same as $\mathbf{\Theta^{i}}$, and there is no need to continue the Viterbi training, since all future models will remain the same.

1. Consider EM for HMMs (Baum-Welch training as outlined on slides 32 and 49. It also has the property that if a parameter in the initial model $\mathbf{\Theta^{0}}$ is set to zero, i.e. if a particular transition or emission probability is set to zero, then it will remain zero during all the iterations of the EM training. Why?

   Solution:

   If a particular transition or emission probability is set to zero in $\mathbf{\Theta^{0}}$, then this transition or emission cannot be part of the any explanation of our traning data $\mathbf{X}$, i.e. the expected number of times the transition or emission is observed will be zero. This means that the transition or emission probability remains zero in new model computed using the formulas on slide 47, and similarly remains zero in all future models $\mathbf{\Theta^{i}}$ that will be generated in the iterative EM-training proces.

Exercise 2: Questions to slides *Hidden Markov Models - Selecting the initial model parameters and using HMMs for (simpel) gene finding*:

1. Consider the 7-state HMM on slides 26 that you also use in pratical exercises. As stated on slide 27, this HMM is also relevant for gene finding, where we say that state 3 emits non-coding symbols, states 2, 1, 0 emit coding triplets (codons) in the left-to-right direction and states 4, 5, 6 emit coding symbols in the reverse (right-to-left) direction.

   If we are given a DNA string, say

   `ACGTATGCTAATCTAAACCTACGGCATGT`

   and information about its gene structure using the N, C, R annotation also used in the slides and practical exercises, say

   `NNNNCCCCCCCCCCCCNNRRRRRRRRRNN`

   then we can convert this gene structure into an actual sequence of states, as also explained on slide 30 (for a different model), as

   `333321021021021033456456456 33`

   Use the above DNA string and information about its gene structure to set the model parameters of the 7-state HMM using Traning-by-Counting. (You can perhaps use this small example as a test case for your implementation of

Traning-by-Counting in the practical exercises.)

Solution:

Counting and setting start probabilities:

We only see one start in state 3, so

P(start in 3) = 1

Counting and setting transition probabilities:

From the transition diagram on slide 26, we already know that some transition probability should be 0 and 1. We 'only' need to estimate the probability of the transitions: 3->2, 3->3, 3->4, 0->2, 0->3, 6->4, 6->3. By counting we get:

#(3->2) = 1

#(3->3) = 5

#(3->4) = 1

P(3->2) = 1 / (1+5+1) = 0,143

P(3->3) = 5 / (1+5+1) = 0,713

P(3->4) = 1 / (1+5+1) = 0,143

#(0->2) = 3

#(0->3) = 1

P(0->2) = 3 / (3+1) = 0,75

P(0->3) = 1 / (3+1) = 0,25

#(6->4) = 2

#(6->3) = 1

P(6->4) = 2 / (2+1) = 0,667

P(6->3) = 1 / (2+1) = 0,333

Counting and setting emission probabilities:

For each of the 7 states, we count how many times we see each symbol A, C, G, and T:

State 0: 2 1 1 0

State 1: 1 0 0 3

State 2: 2 1 0 1

State 3: 2 2 2 2

State 4: 0 3 0 0

State 5: 1 0 1 1

State 6: 1 0 1 1

For each state, This translates into the following emission probabilities of A, C, G, and T:

State 0: 0,50 0,25 0,25 0,00

State 1: 0,25 0,00 0,00 0,75

State 2: 0,50 0,25 0,00 0,25

State 3: 0,25 0,25 0,25 0,25

State 4: 0,00 1,00 0,00 0,00

State 5: 0,33 0,00 0,33 0,34

State 6: 0,33 0,00 0,33 0,34

In [ ]: