Assignment 2

Dmitry Adamskiy

December 1, 2020

This assignment is focused on the inference and learning in Belief Networks.

0. **Fully observed case. Weather station data**. You have the data from a weather station (`./meteo0.csv`) recording the weather conditions on consecutive days. You have $N = 500$ sequences of length $T = 100$ each.

   Each observation v_t^n shows whether it was Rainy(0), Cloudy(1) or Sunny(2) on a given day, $v_t^n \in \{0, 1, 2\}$.

   You model the data as first-order Markov chain with transition probabilities A_{ij}, $A_{ij} = P(v_{t+1} = i | v_t = j)$.

   (a) What are the parameters of this model? Find the maximum likelihood parameters of the model. [10 marks]

1. **Three weather stations**. Now you have the data from three weather stations (`meteo1.csv`) and you don't know which data sequences come from which one. You believe that the the data for each weather station follow a first-order Markov chain.

   (a) Draw the graphical model for this problem. What are the parameters of this model? Describe all the steps of the EM-algorithm in this scenario. [10 marks]

   (b) Implement the EM algorithm for the data provided. Print the learned parameters and the log-likelihood of the data for those parameters. For the first 10 rows of the dataset print the posterior distribution of which stations these data sequences come from. [10 marks]

   (c) Explain whether the learned parameters depend on the initial guess for the parameters of the model. Explain your initialisation strategy. Describe any computational issues you encountered when implementing EM and how you solved them. [10 marks]

   (d) Suppose that you don't have any prior knowledge about the parameters. Your friend tells you that you should then initialise all the relevant ditributions to be uniform. Is this a good idea? What would the parameters learned by EM look like and why? [10 marks]

2. **Chess results.** Anton and Dmitry like to play blitz chess games online. While Dmitry always plays in the same style, Anton has two different ways of approaching chess. Anton in his normal mood is a cautious and strong player. However, occasionally he has streaks of bad mood when he plays more agressively, never accepts draws and plays every game till the decisive result[1].

   You have a sequence of outcomes from the games they played. You model it as Hidden Markov Model, where the hidden state at time t is a binary variable indicating whether Anton is in his normal mood ($h_t = 0$) or a bad mood ($h_t = 1$) and the observed variable at time t is the outcome of the game: win for Anton ($v_t = 0$), draw ($v_t = 1$) or win for Dmitry ($v_t = 2$).

   The emission probabilities are known: $P(v_t|h_t = 0) = [0.3, 0.6, 0.1]$ and $P(v_t|h_t = 1) = [0.5, 0, 0.5]$. That is, in the "normal" situation a draw is the most likely outcome, followed by a victory for Anton, whereas in the "bad mood" situation a win for each player is equally likely and there are no draws. The initial hidden state is also known $P(h_1 = 0) = 1$.

   However the transition probabilities $T_{ij} = P(h_{t+1} = i|h_t = j)$ that govern the dynamics of Anton's mood are not known and you have to learn them from the data.

   (a) When the sequence of outcomes $v_{1:T}$ is observed, they induce a distribution of hidden variables $P_v(h_{1:T}) = P(h_{1:T}|v_{1:T})$. What is the graph representation of this distribution? Based on the graphical model, is h_1 independent of h_T in this distribution? [5 marks]

   (b) Suppose that you observe a sequence $v_{1:10} = [0, 1, 0, 2, 0, 2, 1, 0, 2, 0]$. Is h_1 independent of h_{10} given this particular sequence? Why is that the case? [5 marks]

   (c) Implement the forward-backward algorithm to compute the singleton marginals $P(h_t|v_{1:T})$. Compute and report them for the sequence $v_{1:10} = [0, 1, 0, 2, 0, 2, 1, 0, 2, 0]$ and $T = \begin{pmatrix} 0.5 & 0.8 \\ 0.5 & 0.2 \end{pmatrix}$. [10 marks].

   (d) Use the messages computed in the previous part to compute the pairwise marginals $P(h_t, h_{t+1}|v_{1:T})$. Report $P(h_1, h_2|v_{1:10})$ and $P(h_4, h_5|v_{1:10})$ for the same observed sequence and transition matrix as above. [10 marks]

   (e) Implement the EM (Baum-Welch) algorithm. Describe any computational issues you encountered when implementing EM and how you solved them. Run the EM alorithm for the observed sequence $v_{1:1000}$ in the file `chess.csv` starting from the initial guess $T_0 = \begin{pmatrix} 0.5 & 0.8 \\ 0.5 & 0.2 \end{pmatrix}$. Report the number of iterations that it took your implementation to converge and the value of the learned matrix T. [20 marks]

   (f) Using the learned matrix T, predict the outcome of the next game, $P(v_{1001}|v_{1:1000})$. [Bonus 10 marks]

[1] In this berserk state if he gets a drawn position, he resigns or loses on time.