Assignment 1

Dmitry Adamskiy

October 26, 2019

This assignment is focused on the basics of probability theory and Bayesian inference.

1. **Three Cards** One card is white on both faces; one is black on both faces; and one is white on one side and black on the other. The three cards are shuffled and their orientations randomized. One card is drawn and placed on the table. The upper face is black. What is the colour of its lower face? (Solve the inference problem.)

   (this exercise is from David J.C. MacKay's book) [15 marks]

2. Solve the following exercises from the textbook:

   (a) **Earthquakes-1** Exercise 1.22 [5 marks for each subquestion, 20 in total]

   (b) **Earthquakes-2** Exercise 1.23 [5 marks for each subquestion, 15 in total]

3. **Meeting scheduling** You are organising a trip to Scotland for your N friends. You booked the tickets on the Caledonian Sleeper train, departing from Euston station. Since you have all the tickets you decide that you need to meet at Euston station meeting point before the ticket gates. You all need to be there at 21:05 in order not to miss the train. However, it makes sense to ask people to come a bit earlier in case some of them are delayed. But how much earlier?

   (a) Here is the model for the delays that you are going to use. You are always on time. Let D_i be a delay of your i-th friend. You assume that all D_i-s are independent and identically distributed. $P(D_i \leq 0) = 0.7$ (that is, with probability 0.7 your friend will come on time or earlier), $P(0 < D_i < 5 \text{ mins}) = 0.1$, $P(5 \text{ mins} \leq D_i < 10 \text{ mins}) = 0.1$, $P(10 \text{ mins} \leq D_i < 15 \text{ mins}) = 0.07$, $P(15 \text{ mins} \leq D_i < 20 \text{ mins}) = 0.02$, $P(20 \text{ mins} \leq D_i) = 0.01$. You would like to meet as late as possible but still still catch a train with probability at least 0.9. What time $T_0 = T_0(N)$ should you ask your friends to meet? Solve for $N = 3$, $N = 5$, $N = 10$. [10 marks]

   (b) You realise that some people are less punctual than others. You update your model with the unobserved binary variables Z_i. The probabilities $P(D_i|Z_i = \text{punctual})$ are the same as above and $P(D_i|Z_i =$

not punctual) $= (0.5, 0.2, 0.1, 0.1, 0.05, 0.05)$ where the states are ordered as above. You have a prior belief that $p(Z_i = \text{punctual}) = 2/3$ for all i. What are the probabilities of missing the train if you use the answers from (a) for this model? [10 marks]

Bonus Suppose that for $N = 5$ you used the answer from (a) and missed the train. Now you are wondering how many of your N friends are not punctual. What is the posterior distribution of this count? [10 marks]

4. **Dunwich Hamlet** A young and purely fictitious football team Dunwich Hamlet [1] has two types of fans. One type is a hardcore football aficionado, who buys season tickets and rarely misses games. The other type is a typical resident of Dunwich village, who supports his local team but does not go to the games very often himself. Dunwich is a family-friendly club and each ticket holder is allowed to bring a family member for free.

   Now the first season comes to an end and the club chairman would like to thank season ticket holders and invite them to a party. But alas the club accounting is not very mature: the database on season ticket sales was lost in an unfortunate accident[2] and no one has ever counted tickets sold on the day.

   However, the chairman holds every copy of Non League Paper which covers Dunwich games and the paper accurately reports the number of spectators for each game. Can he get a good estimate of how many season ticket holders there are from these data?

   The model and questions. There are a season ticket holders and b "normal" fans. For each game each fan independently decides whether to go to this game or not. The probability for the season ticket holder to attend the game is p_a and for the "normal" fan it is p_b. The prior distributions on a and b are uniform distributions on $[a_{min}, a_{max}]$ and $[b_{min}, b_{max}]$ respectively. Let c_n be the total number of ticket holders for the n-th game.

   (a) Describe the distribution of the number of ticket holders attending the n-th game $p(c_n|a, b)$. What are its mean and variance? [2 marks]

   (b) Each ticket holder (independently) with probability p_d brings a family member. Let d_n be the total number of fans at n-th game (this is the number reported in the paper). What is the mean of the distribution $p(d_n|c_n)$? [3 marks]

   (c) Draw the graphical model corresponding to this setup (the variables are $a, b, c_1, c_2, \ldots, c_N, d_1, d_2, \ldots, d_N$). [5 marks]

   (d) Write a program that computes posterior distributions $P(a|d_1, d_2, \ldots, d_n)$ and $P(b|d_1, d_2, \ldots, d_n)$ and compute these posteriors for $n = 1, 2, \ldots, 10$. Plot these posteriors. What are the ML/MAP estimates of a and b after taking all ten games into account? Use the following model parameters: $a_{min} = 0, a_{max} = 15, b_{min} = 0, b_{max} = 200, p_a =$

[1] Not to be confused with Dulwich Hamlet or Dunwich Dynamo.

[2] Dog ate chairman's notebook

$0.99, p_b = 0.3, p_d = 0.5$. The data sequence $D = [d_1, \dots, d_{10}]$ is $[22, 27, 26, 32, 31, 25, 35, 26, 28, 23]$. [20 marks]