Exercise Session 6

1 Beta prior

We are doing a statistical experiment in order to estimate the probability of head (H) and tail (T) of a coin. We flip the coin 10 times obtaining the sequence:

$$D = \{H, T, H, T, T, T, H, T, T, T\} \tag{1}$$

a) What is the maximum likelihood estimate θ_{MLE} of $p(H)$?

b) What is the Bayes estimate θ_{Bayes} of $p(H)$ using a prior $Beta(\theta; 0.0001, 0.0001)$?

Expected value of the posterior distribution

c) What is the Bayes estimate θ_{Bayes} of $p(H)$ using a prior $Beta(\theta; 10000, 10000)$?

2 Maximum à posteriori (MAP) estimation with a Beta prior

In the previous exercise we calculated the Bayes estimate when given some observations and a Beta prior. It was obtained by calculating the mean of the resulting posterior distribution $Beta(\theta; \alpha, \beta)$: $\langle\theta\rangle = \alpha/(\alpha+\beta)$. In this exercise we consider the same data D as in the previous exercise but we are going to calculate the MAP estimate. We obtain it not through the mean of the posterior distribution but through its mode. (Look in up the corresponding formula in the lecture notes.)

The mode of the posterior distribution

a) What is the MAP estimate θ_{MAP} for $p(H)$ when no prior information is available?

b) What is the MAP estimate θ_{MAP} for $p(H)$ using a prior $Beta(\theta; 0.0001, 0.0001)$?

c) What is the MAP estimate θ_{MAP} for $p(H)$ using a prior $Beta(\theta; 10000, 10000)$?

3 Missing at random

The systolic blood pressure of 50 people is measured. Each person is measured twice (one month apart). Measurements are positively correlated. We simulate three possible causes for missing values:

a) There was a problem with the hard drive and no backup is available. Data were corrupted and some values are missing.

b) Everybody is tested twice, but the second value is recorded only if higher than 140.

c) Only people with pressure higher than 140 are invited to the second test.

Draw the graphical model for each case and explain which assumption between MAR, MCAR, MNAR holds. Use t_1, t_2 as test variables and m_1 m_2 as missingness variables.

4 Expectation Maximization

Consider the Bayesian network $p(a, b) = p(a)p(b|a)$ (Figure 1), where the variables are binary.

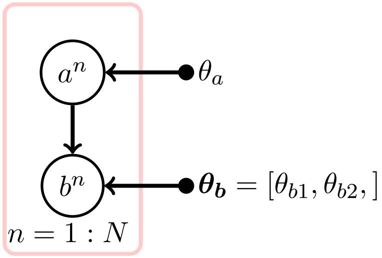

Figure 1: Bayesian network

The parameters $\boldsymbol{\theta}$ are defined as: $p(a = 1) = \theta_a$, $p(b = 1|a = 1) = \theta_{b1}$, and $p(b = 1|a = 0) = \theta_{b2}$. We observe the data in Table 1. The question marks indicate the fact that some data is missing (at random).

a	b
1	1
1	?
0	0
?	0

Table 1: Database of observations. The question marks indicate missing data.

In this exercise we will use classical Expectation Maximization in order to learn the parameters of the conditional probability tables $p(a)$ and $p(b|a)$. *(hint: use the procedure explained in 11.2.4 in the textbook. Beware, the equation in line 9 of Algorithm 11.2 should be replaced by equation 11.2.43 (this is correct in the latest online pdf).)*

- Provide the equations for the E and M steps that will be used in the expectation maximisation algorithm. Assume that you have current parameter estimates θ_a^{old}, θ_{b1}^{old}, and θ_{b2}^{old} available.
- Perform two iterations of the EM algorithm starting from the initial estimates: $\theta_a = 0.4$, $\theta_{b1} = 0.2$, and $\theta_{b2} = 0.3$.

5 Yet another EM exercise

Two doctors have been independently using two different tests to diagnose whether a patient has the flu or not. At some point, the two doctors meet to figure out which of the tests has the higher reliability. Some of the patients did not agree to be tested twice, and some patients did not follow-up to confirm whether they actually had the flu or not. Nevertheless, the doctors do not want to waste any medical record, even if some of the information is missing. The data they collected is shown in Table 2.

a) Draw the belief network that you would use to model this problem.

b) Assuming that the prior probabilities:

- $p(\text{outcome test 1} = \textit{neg}|\text{patient has the flu} = \textit{no}) = 0.80$
- $p(\text{outcome test 1} = \textit{neg}|\text{patient has the flu} = \textit{yes}) = 0.15$
- $p(\text{outcome test 2} = \textit{neg}|\text{patient has the flu} = \textit{no}) = 0.9$
- $p(\text{outcome test 2} = \textit{neg}|\text{patient has the flu} = \textit{yes}) = 0.25$

patient index	test 1 outcome	test2 outcome	really has the flu?
1	positive	?	yes
2	positive	positive	yes
3	?	negative	no
4	positive	negative	yes
5	?	positive	?

Table 2: Patient data for Exercise 5.

- $p(\text{patient has the flu} = yes) = 0.50$

Describe in detail the updates for E and M steps for patient 1 and for at least 1 CPT.

c) What is the value of the parameters:

- $p(\text{outcome test 1} = neg|\text{patient has the flu} = no)$
- $p(\text{outcome test 1} = neg|\text{patient has the flu} = yes)$

after one full iteration of the EM algorithm using the whole dataset

1) $\theta_{MLE} = \frac{3}{10} = 0.3$

2) $\theta \sim B(\theta; 0.0001, 0.0001)$

$$\theta_{Bayes} = \frac{3.0001}{10.0002} = 0.3$$

$$\theta_{Bayes} = \frac{10003}{20010} \approx 0.4999$$