Probabilistic Reasoning in AI - Assignment 4

Due Wednesday, March 12

1. [45 points] **Normal distribution**

   The normal distribution is one of the most important probability distributions, used widely in many scientific disciplines. This exercise is meant as a review of some of the basic properties for the normal distribution. It is useful as an exercise to compute these answers by yourself rather than looking them up.

   (a) [10 points] The univariate normal distribution has the following probability density function:

   $$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

   where μ and σ are the mean and variance parameters. Suppose that you do not know μ and σ, but you have examples $x_1, \ldots x_n$ drawn i.i.d. from the distribution in which you are interested. Write down the log likelihood of the data, take the derivatives with respect to μ and σ, set them to 0 and derive the expression of the maximum likelihood estimators for these parameters, denoted $\hat{\mu}^n$ and $\hat{\sigma}_n$.

   (b) [5 points] Suppose that you now get a new data point x_{n+1}. Write down an update rule which computes μ_{n+1} and σ_{n+1} from μ_n, σ_n and x_{n+1} (without involving any of the previous data points $x_1, \ldots x_n$)

   (c) [10 points] Suppose that you know the variance parameter σ but the mean μ is unknown. Instead, you have a *prior distribution* for μ, which is also normal, with mean μ_0 and variance σ_0. Show that the *posterior* distribution of the mean, after seeing data $x_1, \ldots x_n$ is also a normal distribution, and give analytical expressions for its mean and variance. In other words, the normal distribution is a conjugate prior for itself. What happens with the posterior mean distribution as the number of data points $n \to \infty$?

   (d) [10 points] In Matlab, draw a set of 10 data points from a Normal distribution of mean 0 and variance 1. Suppose that you have a prior for the mean, which is of mean 1 and variance 2. Use the formula above to compute the new mean distribution. Do the same with 100 real data points. Plot on the same graph the two mean distributions and the true mean, and explain what you see.

   (e) [10 points] Now suppose that you do not use the prior directly, but instead draw an equivalent sample size of k points from it. These points will get mixed up with the data coming from the true distribution. For two equivalent sample sizes, consisting of 10 points and 100 points respectively, repeat the exercise above. Explain how the equivalent sample size affect the results you observe.

2. [25 points] **Expectation maximization**

   Suppose that somebody gave you a bag with two biased coins, having the probability of coming up heads of p_1 and p_2 respectively. You are supposed to figure out p_1 and p_2 by tossing the coins repeatedly. You will repeat the following experiment n times: pick a coin uniformly at random from the bag (i.e. each coin has probability $1/2$ of being picked) and toss it, recording the outcome (heads or tails). The coin is then returned to the bag. Assume that the individual experiments are independent.

(a) [5 points] Set up a graphical model which captures this problem.

(b) [5 points] Suppose the two coins have different color: the p_1 coin is white, the other coin is yellow. Show the maximum likelihood estimators for the two parameters, p_1 and p_2.

(c) [5 points] Suppose now that the two coins look identical, so when you take them out of the bag you cannot tell them apart. Hence, the identity of the coin is always missing in your data. Write the expected log-likelihood of the data in this case.

(d) [5 points] Suppose you start with some guesses for the parameters, $\hat{p_1}$ and $\hat{p_2}$. Show the E-step of the soft EM algorithm.

(e) [5 points] Show the M-step of the EM algorithm.

3. [30 points] **Structure learning in Bayes nets**

In this problem, we consider the task of learning the structure of a Bayes net, in the special case in which we require this structure to be a tree. We would like to find the tree structure with maximum likelihood.

(a) [10 points] Consider an undirected tree T over r.v.s $X_1, \ldots X_n$. Show that all Bayes nets that share T as a skeleton have the same likelihood score. Derive the formula for this score.

(b) [20 points] Based on this formula, write an algorithm which will search for the tree with the best score. Hint: it will be useful to use the maximum spanning tree algorithm.