DD2434 Machine Learning, Advanced Course
Assignment 2

Jens Lagergren

Deadline 12.00 (noon) (CET) December 22th, 2016

You will present the assignment will by a written report that you can mail to me at

jens.lagergren@scilifelab.se

before the deadline. From the report it should be clear what you have done and you need to support your claims with results. You are supposed to write down the answers to the specific questions detailed for each task. This report should clearly show how you have drawn your conclusions and explain your derivations. Your assumptions, if any, should be stated clearly. For the practical part of the task you should not show any of your code (but we will ask randomly selected and potentially also other students to provide their code) but rather only show the results of your experiments using images and graphs together with your analysis.

Being able to communicate your results and conclusions is a key aspect of any scientific practitioner. It is up to you as a author to make sure that the report clearly shows what you have done. Based on this, and only this, we will decide if you pass the task. No detective work should be needed on our side. Therefore, neat and tidy reports please!

The grading of the assignment will be as follows,

- **E** Completed Tasks 2.1, 2.2, and 2.3.
- **D** E + Completed one of Tasks 2.4, 2.6, 2.5, and 2.7.
- **C** E + Completed two of Tasks 2.4, 2.6, 2.5, and 2.7.
- **B** E + Completed three of Tasks 2.4, 2.6, 2.5, and 2.7.
- **A** Completed all tasks.

These grades are valid for review December 18th, 2015. Delayed assignments can only receive the grade E.

I Graphical Models

2.1 Qualitiative effects in a Directed Graphical Model (DGM)

Consider the Directed Acyclical Graph (DAG) of a DGM shown in Figure 1. The variables are binary-valued. The conditional probability densities are not known but, in contrast, available information reveal how each variable qualitatively influences its children. The interpretation of the influences, which are denoted $\overset{+}{\to}$ and $\overset{-}{\to}$, are:

$\overset{+}{\to}$ means $p(y^1|x^1, \mathbf{c}) > p(y^1|x^0, \mathbf{c})$, for all values $\mathbf{c}$ of Y's parents.

$\overset{-}{\to}$ means $p(y^1|x^1, \mathbf{c}) < p(y^1|x^0, \mathbf{c})$, for all values $\mathbf{c}$ of Y's parents.

You should also assume the parents in any V-structure are conditionally dependent given the the common child. Assume that no probability involved equals 0 or 1. Consider the following pairs of conditional probabilities:

1. $p(t^1|d^1)$ and $p(t^1)$
2. $p(d^1|t^0)$ and $p(d^1)$
3. $p(h^1|e^1, f^1)$ and $p(h^1|e^1)$
4. $p(c^1|f^0)$ and $p(c^1)$
5. $p(c^1|h^0)$ and $p(c^1)$
6. $p(c^1|h^0, f^0)$ and $p(c^1|h^0)$
7. $p(d^1|h^1, e^0)$ and $p(d^1|h^1)$
8. $p(d^1|e^1, f^0, w^1)$ and $p(d^1|e^1, f^0)$
9. $p(t^1|w^1, f^0)$ and $p(t^1|w^1)$.

Answer the following questions (in this problem you do not have to motivate your answers).

Question 1: *In which pairs is one value larger than the other?*

Question 2: *Which pairs are equal?*

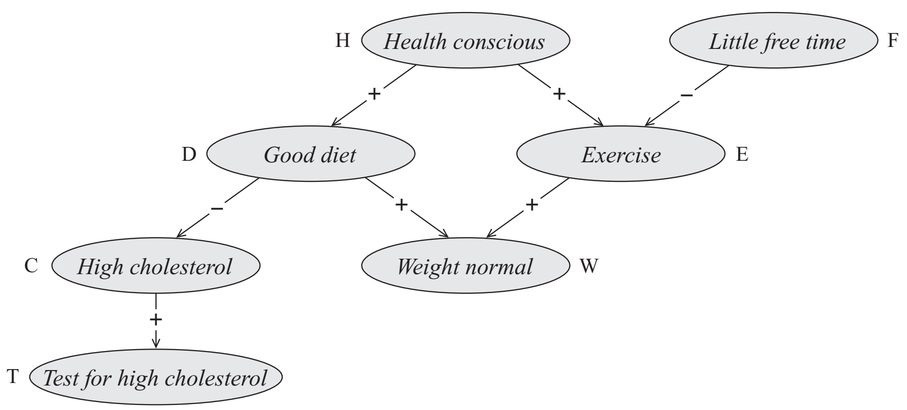

Figure 1: The DAG with qualitative influence information.

Question 3: *Which pairs are incomparable (i.e., the two values can not be compared based on the information available in the DAG.)*

2.2 The casino model

Consider the following generative model. There are $2K$ tables in a casino, $t_1, ..., t_K, t'_1, ..., t'_K$ of which each is equipped with a single dice (which may be biased, i.e., any categorical distribution on $\{1, ..., 6\}$) and N players $P_1, ..., P_N$ of which each is equipped with a single dice (which also may be biased, i.e., any categorical distribution on $\{1, ..., 6\}$). Each player P_i visits K tables. In the k:th step, if the previous table visited was t_{k-1}, the player visits t_k with probability $1/4$ and t'_k with probability $3/4$, and if the previous table visited was t'_{k-1}, the player visits t'_k with probability $1/4$ and t_k with probability $3/4$. So, in each step the probability of staying among the primed or unprimed tables is $1/4$. At table k player i throws her own dice as well as the table's dice. We then observe the sum S^i_k of the two dice, while the outcome of the table's dice X_k and the player's dice Z_k are hidden variables. So for player i, we observe $S^i = S^i_1, ..., S^i_K$, and the overall observation for N players is $S^1, ..., S^N$.

Question 4: *Provide a drawing of the Casino model as a graphical model. It should have a variable indicating the table visited in the k:th step, variables for all the dice outcomes, variables for the sums, and plate notation should be used to clarify that N players are involved.*

Question 5: *Implement the Casino model (in Matlab or Python).*

Question 6: *Provide data generated using at least three different sets of categorical dice distributions – what does it look like for all unbiased dice, i.e., uniform distributions, for example, or if some are biased in the same way, or if some are unbiased and there are two different groups of biased dice*

2.3 Simple VI

Consider the model defined by Equation (10.21)-(10-23) in Bishop. We are here concerned with the VI algorithm for this model covered during the lectures and in the book.

Question 7: *Implement the VI algorithm for the variational distribution in Equation (10.24) in Bishop.*

Question 8: *Describe the exact posterior*

Question 9: *Compare the variational distribution with the exact posterior. Run the inference for a couple of interesting cases and describe the difference.*

2.4 Sampling tables given dice sums

You will now design an algorithm that does inference on the casino model that you designed in Task 2.2.

Question 10: *Describe an algorithm that, given (1) the parameters* Θ *of the full casino model of Task 2.2 (so,* Θ *is all the categorical distributions corresponding to all the dice), (2) a sequence of tables* $r_1 \ldots, r_n$ *(that is,* r_i *is* t_i *or* t'_i*) , and (3) an observation of dice sums* $s_1, \ldots, s_K$*, outputs* $p(r_1, ..., r_K | s_1, \ldots, s_K, \Theta)$.

Notice, in the DP algorithm for the above problem you have to keep track of the last table visited.

Question 11: *You should also show how to sample* $r_1, \ldots, r_K$ *from* $p(R_1, ..., R_K | s_1, \ldots, s_K, \Theta)$ *as well as implement and show test runs of this algorithm. In order to design this algorithm show first how to sample* r_K *from*

$$p(R_K | s_1, \ldots, s_K, \Theta) = p(R_K, s_1, \ldots, s_K | \Theta) / p(s_1, \ldots, s_K | \Theta)$$

and then r_{K-1} *from*

$$p(R_{K-1} | r_K, s_1, \ldots, s_K, \Theta) = p(R_{K-1}, r_K, s_1, \ldots, s_K | \Theta) / p(r_K, s_1, \ldots, s_K | \Theta).$$

2.5 Expectation-Maximization (EM)

Consider the following simplification of the casino model from Problem 2.2. There are K tables in the casino $t_1, ..., t_K$ of which each is equipped with a single dice (which may be biased, i.e., any categorical distribution on $\{1, ..., 6\}$) and N players $P_1, ..., P_N$ of which each is equipped with a single dice (which also may be biased, i.e., any categorical distribution on $\{1, ..., 6\}$). Let Θ be the parameters of all these categorical distributions.

Each player P_i visits the K tables in the order $1, ..., K$. At table k the player i throws her own dice as well as the table's dice. We then observe the sum S_k^i of the dice, while the outcome of the tables dice X_k and the player's dice Z_k are hidden variables. So for player i, we observe $s^i = s_1^i, ..., s_K^i$, and the overall observation for N players is $s^1, ..., s^N$.

Design and describe an EM algorithm for this model. That is, an EM algorithm that given $s^1, ..., s^N$ finds locally optimal parameters for the categorical distributions (i.e., the dice), that is, the Θ maximising $P(s_1^i, ..., s_K^i | \Theta)$.

Question 12: *Present the algorithm written down in a formal manner (using both text and mathematical notation, but not pseudo code).*

Question 13: *Implement it and test the implementation with data generated in Task 2.2, and provide graphs or tables of the results of testing it with the data.*

2.6 Variational Inference

Again consider a variation of the of the casino model used in the EM problem. There are again K tables in the casino $t_1, \ldots, t_K$ and N players $P_1, \ldots, P_N$, but in the present case they are all equipped with Gaussian distributions. In fact, t_k is equipped with $N(X | \mu_k, \lambda_k^{-1})$, where μ_k has prior distribution $N(Y | \mu, \lambda^{-1})$ and λ_k is known, and P_n is equipped with $N(Z | \xi_k, \iota_k^{-1})$, where ξ_k has prior distribution $N(X_k | \xi, \iota^{-1})$ and ι is known. As before, each player visits each table in the order $t_1, \ldots, t_K$ and, when P_n visits table k, we observe $S_{ik} = X + Y$ where X is sampled from the tables Gaussian and Y from the players Gaussian. Let the data be $D = s_1, \ldots, s_K$ where $s_n = s_{n1}, \ldots, s_{nK}$ are the sums obtained for player n.

Use Variational Inference in order to obtained a variational distribution

$$q(\mu_1, \ldots, \mu_K, \xi_1, \ldots, \xi_N) = \prod_k q(\mu_k) \prod_n q(\xi_n)$$

that approximates $p(\mu_1, \ldots, \mu_K, \xi_1, \ldots, \xi_N | D)$.

2.7 Harder VI

Now let's take the model from the previous problem and reintroduce tables pairs. Now there $2K$ tables in the casino $t_1, \ldots, t_K$ and $t'_1, \ldots, t'_K$ as well as N players $P_1, \ldots, P_N$, also in the present case they are all equipped with Gaussian distributions. As above, t_k is equipped with $N(X|\mu_k, \lambda_k^{-1})$, where μ_k has prior distribution $N(Y|\mu, \lambda^{-1})$ and λ_k is known, and P_n is equipped with $N(Z|\xi_k, \iota_k^{-1})$, where ξ_k has prior distribution $N(X_k|\xi, \iota^{-1})$ and ι is known. In addition, t'_k is equipped with $N(X|\kappa_k, \nu_k^{-1})$, where κ_k has prior distribution $N(Y|\kappa, \nu^{-1})$ and ν_k is known.

As previously been the case for table pairs each player P_n visits K tables as follows. In the k:th step, if the the previous table visited was t_{k-1}, the player visits t_k with probability $1/4$ and t'_k with probability $3/4$, and if the previous table visited was t'_{k-1}, the player visits t'_k with probability $1/4$ and t_k with probability $3/4$. So, in each step the probability of staying among the primed or unprimed tables is $1/4$. When P_n visits table k, we observe $S_{ik} = X + Y$ where X is sampled from the tables Gaussian and Y from the players Gaussian. Let the data be $D = s_1, \ldots, s_K$ where $s_n = s_{n1}, \ldots, s_{nK}$ are the sums obtained for player n.

Use Variational Inference in order to obtained a variational distribution

$$q(\mu_1, \ldots, \mu_K, \xi_1, \ldots, \xi_N) = \prod_k q(\mu_k) \prod_n q(\xi_n)$$

that approximates $p(\mu_1, \ldots, \mu_K, \xi_1, \ldots, \xi_N | D)$.

Hint: first express $p(s_n|\mu_1, \ldots, \mu_K, \kappa_1, \ldots, \kappa_K, \xi_1, \ldots, \xi_N)$ as a sum over all feasible sequences of tables, i.e.,

$$\sum_{\substack{W_1, \ldots, W_K \\ W_k \in \{t_k, t'_k\}}} \prod_{k=1}^{K} p(W_1, \ldots, W_K) p(s_n | \mu_1, \ldots, \mu_K, \kappa_1, \ldots, \kappa_K, \xi_1, \ldots, \xi_N \xi_1, \ldots, \xi_N, W_1, \ldots, W_K).$$

Good Luck!