DD2434 Machine Learning, Advanced Course
Assignment 2

Jens Lagergren

Deadline 23.00 (CET) December 30, 2017

You will present the assignment by a written report, submitted before the deadline using Canvas . You must solve the assignment individually and it will automatically be checked for similarities to other students' solutions as well as document on the web in general. 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. Show the results of your experiments using images and graphs together with your analysis and add your code as an appendix.

Being able to communicate 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 required on our side. In particular, neat and tidy reports please!

The grading of the assignment will be as follows,

E Correctly completed Task 2.1, 2.2, and 2.3.

D E + Correctly completed one of Task 2.4, 2.5, 2.6, and 2.7.

C E + Correctly completed two of Task 2.4, 2.5, 2.6, and 2.7.

B E + Correctly completed three of Task 2.4, 2.5, 2.6, and 2.7.

A Correctly completed all tasks.

These grades are valid for assignments submitted before the deadline, late assignments can at most receive the grade E.

Good Luck!

I Graphical Models

2.1 Dependences in a Directed Graphical Model

Consider the Directed Acyclic Graph (DAG) of a DGM shown in Figure 1.

Question 1: *Which pairs of variables, not including X, are dependent conditioned on X?*

Question 2: *Which pairs of variables, not including X, are dependent, not conditioned on X?*

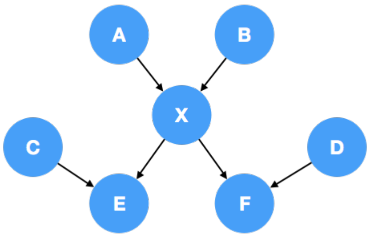

Figure 1: The DAG

2.2 The Sum-HMM

Consider the following generative model, which is based on a standard HMM. There are $2K$ tables in a casino, $t_1^1, ..., t_K^1, t_1^2, ..., t_K^2$ of which each is equipped with a single dice which may be biased, i.e., any categorical distribution $p(X_k|Z_k = t_k^i)$ on $\{1, ..., 6\}$. There are N players $p_1, ..., p_N$. Each player p_n visits K tables. In the k:th step, if the previous table visited was t_{k-1}^i, the player visits t_k^i with probability $1/4$ and t_k^{3-i} with probability $3/4$. So, in each step the probability of staying among the tables in the first group, $t_1^1, ..., t_K^1$, or the second group, $t_1^2, ..., t_K^2$, is $1/4$. At table k player n throws the table's dice; its outcome X_k^n is with probability p (p is known to us) observed and with probability $1 - p$ hidden. We finally observe the sum $S^n = \sum_{k=1}^{K} X_k^n$ of all the outcomes for the player. So, the overall observation for N players is $S^1, X^1, ..., S^N X^N$, where X^n is the sequence of outcomes for player n with the unobserved outcomes censored (say each such entry is "?").

Question 3: *Implement the Sum-HMM, i.e., write your own code for it.*

Question 4: *Provide data generated using at least three different sets of categorical dice distributions that provide reasonable tests for the correctness of your program .*

Question 5: *Motivate your test and why the result of it indicates correctness.*

Question 6: *Give polynomial time dynamic programming algorithm for computing $p(X_k^n = s, Z_k = t_k^i | s^n, x^n)$. Hint: a dice outcome is an integer between 1 and 6, so a sum s^n is an integer between K and $6K$ and, moreover, if a partial sum is associated with a state t_k^i, it is an integer between k and $6k$.*

Question 7: *Implement this DP algorithm, test it, in particular for varying values of p, and, finally, motivate your tests and why the result of it indicates correctness.*

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

Question 9: *Describe the exact posterior*

Question 10: *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 for the Sum-HMM

You will now design an algorithm that does inference on the Sum-HMM model from Task 2.2.

Question 11: *Describe an algorithm that, given (1) the parameters Θ of the Sum-HMM model of Task 2.2 (so, Θ is p and all the categorical distributions corresponding to all the dice), (2) a sequence of tables $z_1, \ldots, z_K$ (where $z_k \in \{t_k^1, t_k^2\}$), and (3) a single player sum and outcome sequence s^n, x^n, outputs $p(z_1, \ldots, z_K | s^n, x^n, \Theta)$.*

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

Question 12: *You should also show how to sample $Z_1, \ldots, Z_K$ from $p(Z_1, \ldots, Z_K | s^n, x^n, \Theta)$ as well as implement and show test runs of this algorithm. In order to design this algorithm show first how to sample Z_K from*

$$p(Z_K | s^n, x^n, \Theta) = p(Z_K, s^n, x^n | \Theta)/p(s|\Theta)$$

and then Z_{K-1} from

$$p(Z_{K-1} | Z_K, s^n, x^n, \Theta) = p(Z_{K-1}, Z_K, s^n, x^n | \Theta)/p(Z_K, s^n, x^n | \Theta).$$

2.5 Expectation-Maximization (EM)

Consider again the Sum-HMM model from Problem 2.2.

Design and describe an EM algorithm for this model. That is, an EM algorithm that given outputs for N players, $s^1, x^1, \ldots, s^N, x^N$, finds locally optimal parameters for the categorical distributions (i.e., the dice), that is, the Θ maximising $P(s^1, x^1, \ldots, s^N, x^N | \Theta)$. First estimate p, then used it in the EM algorithm.

Question 13: *Present the algorithm written down in a formal manner (using both text and mathematical notation, but not pseudo code).*

Question 14: *Implement the algorithm and test the implementation with data generated in Task 2.2, and provide plots of the results of the test.*

2.6 Variational Inference

The Cartesian Matrix Model (CMM) is defined as follows. There are R row distributions $\{N(\mu_r, \lambda_r^{-1}) : 1 \leq r \leq R\}$, each variance λ_r^{-1} is known and each μ_r has prior distribution $N(\mu, \lambda^{-1})$. There are also C column distributions $\{N(\xi_r, \tau_c^{-1}) : 1 \leq c \leq C\}$, each variance τ_c^{-1} is known and each ξ_c has prior distribution $N(\xi, \tau^{-1})$. All hyper-parameters are known. A matrix S is generated by, for each row $1 \leq r \leq R$ and each column $1 \leq c \leq C$, setting $S_{rc} = X_r + Y_c$ where X_r is sampled from $N(\mu_r, \lambda_r^{-1})$ and Y_c from $N(\xi_r, \tau_c^{-1})$. Use Variational Inference in order to obtained a variational distribution

$$q(\mu_1, \ldots, \mu_R, \xi_1, \ldots, \xi_C) = \prod_r q(\mu_r) \prod_c q(\xi_c)$$

that approximates $p(\mu_1, \ldots, \mu_R, \xi_1, \ldots, \xi_C | S)$. Tip: what distribution do you get from the sum of two Gaussian random variables? What is the relation between the means?

Question 15: *Present the algorithm written down in a formal manner (using both text and mathematical notation, but not pseudo code).*

2.7 Variational Inference

Consider the following casino model. There are again $2K$ tables in a casino, $t_1^1, ..., t_K^1, t_1^2, ..., t_K^2$ and N players $p_1, ..., p_N$. In the present case, however, tables as well as players are equipped with Gaussian distributions. In fact, each table t_k^i is equipped with the gaussian $N(\mu_k^i, \lambda_k^{-1})$ where the variance λ_k^{-1} is known and μ_k^i has prior distribution $N(\mu, \lambda^{-1})$. Similarly, each player p_n is equipped with a Gaussian $N(\xi_n, \iota_n^{-1})$, where the variance is known and ξ_n has prior distribution $N(\xi, \iota^{-1})$. All hyper-parameters are known.

The n:th player visits K tables. As above, in the k:th step, if the previous table visited was t_{k-1}^i, the player visits t_k^i with probability $1/4$ and t_k^{3-i} with probability $3/4$. At table k player n samples X_k^n from the table's Gaussian and Y_k^n from her own Gaussian, we then observe the sum $S_k^n = X_k^n + Y_k^n$, while X_k^n and Y_k^n are hidden. So for player n, we observe $S^n = S_1^n, ..., S_K^n$, and the overall observation for N players is $S^1, ..., S^N$.

Use Variational Inference in order to obtained a variational distribution

$$q(\mu_1^1, \ldots, \mu_K^1, \mu_1^2, \ldots, \mu_K^2, \xi_1, \ldots, \xi_N) = \prod_k q(\mu_k^1) \prod_k q(\mu_k^2) \prod_n q(\xi_n)$$

that approximates $p(\mu_1^1, \ldots, \mu_K^1, \mu_1^2, \ldots, \mu_K^2, \xi_1, \ldots, \xi_N | S^1, ..., S^N)$. First target

$$q(\mu_1^1, \ldots, \mu_K^1, \mu_1^2, \ldots, \mu_K^2, \xi_1, \ldots, \xi_N, Z) = q(Z) \prod_k q(\mu_k^1) \prod_k q(\mu_k^2) \prod_n q(\xi_n),$$

where $Z = Z_1, \ldots, Z_K$ is the table sequence, that approximates

$$p(\mu_1^1, \ldots, \mu_K^1, \mu_1^2, \ldots, \mu_K^2, \xi_1, \ldots, \xi_N, Z | S^1, ..., S^N)$$

and, then, perform the trivial marginalization.

Question 16: *Present the algorithm written down in a formal manner (using both text and mathematical notation, but not pseudo code).*