DD2434 Machine Learning, Advanced Course
Assignment 1: Graphical Models

Jens Lagergren and Hedvig Kjellström

Deadline November 25

The assignment will be reviewed by an oral examination. In the beginning of the review session, the examiner will ask you what grade you aim for, and ask questions related to that grade. All the tasks have to be presented at the same review session you can not complete the assignment with additional tasks after it has been examined and given a grade. Come prepared to the review session! The review will take 15 minutes or less, so have all your results in order. The grading of the assignment will be as follows,

E Completed Task 1.1 and 1.2.

D E + Completed Task 1.3.

C D + Completed Task 1.4.

B C + Completed Task 1.5.

A B + Completed Task 1.6.

These grades are valid for review November 25, 2014. See the course web page, HT 2014 - Assignments in the menu, for grading of delayed assignments.

Abstract

The purpose of this assignment is to gain experience with directed and undirected graphical models, to be able to explain and apply the probabilistic theory, but also to improve and apply problem solving skills in this context, i.e., to modify and extend the treated methods in order to suit different conditions.

1.1 Undirected Graphical Models

Problem A. Let N be an undirected graphical model (UGM) defined by the factors:

$\phi_1(A,B)$			$\phi_2(B,C)$			$\phi_3(C,D)$			$\phi_4(D,A)$		
a^0	b^0	30	b^0	c^0	100	c^0	d^0	1	d^0	a^0	100
a^0	b^1	5	b^0	c^1	1	c^0	d^1	100	d^0	a^1	1
a^1	b^0	1	b^1	c^0	1	c^1	d^0	100	d^1	a^0	1
a^1	b^1	10	b^1	c^1	100	c^1	d^1	1	d^1	a^1	100

What is the probability that a, b, c, and d all equals 0?

Problem B. Explaining away is one type of intercausal reasoning, but other type of intercausal interactions are also possible. Provide a realistic example that exhibits the opposite type of interaction.

Consider a v-structure $X \rightarrow Z \leftarrow Y$ over three binary-valued variables. Construct a conditional probability distribution (CPD) $P(Z|X,Y)$ such that

1. X and Y both increase the probability of the effect, that is, $P(Z=1|X=1) > P(Z=1)$ and $P(Z=1|Y=1) > P(Z=1)$,
2. conditioned by Z, each of X and Y increase the probability of the other, that is, $P(X=1|Z=1) < P(X=1|Y=1,Z=1)$, and similarly $P(Y=1|Z=1) < P(Y=1|X=1,Z=1)$.

Note that strong inequalities should hold in all cases. Your example should be realistic, that is, $X;Y;Z$ should correspond to a real-world example, and the CPD should be reasonable.

At the Review

Present solutions to Problems A and B.

In particular you should for Problem A show a graphical model drawn on paper that represents the same information as the table above, and be able to explain all details on how the different nodes and edges in the graph relate to terms in the equations.

For Problem B you should give account of a real-world example according to the problem statement.

For both problems, show written record of all parts of the derivations, and be prepared to explain all parts of your derivations.

1.2 Generating Data from Directed Graphical Models

Problem D part 1. 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 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 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 the probability of staying on among the primed or unprimed tables is $1/4$. At table k the player i throws the player's 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 own 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 $S^1, ..., S^N$.

At the Review

Show the graphical model Θ in Problem D part 1 drawn on paper.

Show an implementation (in Matlab or Python) of the model Θ, either on paper or on the laptop screen, and be prepared to explain all parts of the code.

Present data generated using at least three different sets of categorical dice distributions – what does it look like for all perfect dice with uniform distributions, for example, or if all of them are perfect instead of one, or if all are bad in the same way?

1.3 Theory for Directed Graphical Models

Problem C. Let G be a Bayesian network structure and H a Markov network structure over X such that the skeleton of G is precisely H. Prove that if G has no immoralities (i.e., no pair of vertices with a common child but no edge between), then $I(G) = I(H)$.

Problem D. Let us revisit the generative model of a casino described in Task 1.2.

Describe a Viterbi-like algorithm that, given the parameters of the model Θ (i.e., all the categorical distributions corresponding to the dice) and an observation $S^i = S^i_1, ..., S^i_K$, outputs the most likely outcomes of the tables dice, i.e., the $x_1, ..., x_K$ maximizing $P(S^i_1, ..., S^i_K | x_1, ..., x_K, \Theta)$.

At the Review

Present complete solutions to Problems C and D.

Present the proof and the algorithm written down in a formal manner, and be prepared to explain each part of them, to show that you understood. Point out possible pitfalls that you have avoided.

1.4 Implementing the Viterbi-like Algorithm for Directed Graphical Models

Implement the Viterbi-like algorithm for Problem D that you developed in Task 1.3. Test it with data generated in Task 1.2.

At the Review

Show an implementation (in Matlab or Python) of the algorithm, either on paper or on the laptop screen, and be prepared to explain all parts of the code.

Show graphs or tables of the results of testing it with the data. Be prepared to explain the results of the evaluation. NOTE: The results should have been prepared beforehand, you do not have to run the code during examination.

1.5 Theory for EM

Problem E. Consider the following simplification of the casino model from Problem D. 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 may be biased, i.e., any categorical distribution on $\{1, ..., 6\}$). Each player P_i visits K tables in the order $1, ..., K$. At table k the player i throws their own dice as well as the tables dice. We then observe the sum S^i_k of the dice, while the outcome of the tables dice X_k and the player's own 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 $S^1, ..., S^N$.

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 $P(S^i_1, ..., S^i_K | \Theta)$.

At the Review

Present complete solutions to Problem E.

Present the algorithm written down in a formal manner, and be prepared to explain each part of it, to show that you understood.

1.6 Implementation of EM

Implement the algorithm for Problem E that you developed in Task 1.5. Test it with data generated in Task 1.2.

At the Review

Show an implementation (in Matlab or Python) of the algorithm, either on paper or on the laptop screen, and be prepared to explain all parts of the code.

Show graphs or tables of the results of testing it with the data. Be prepared to explain the results of the evaluation. NOTE: The results should have been prepared beforehand, you do not have to run the code during examination.

Good Luck!