DD2434 Machine Learning, Advanced Course
Assignment 2

Jens Lagergren

Deadline, see Canvas

Read this before starting

There are several commonalities between the problems and, consequently, it may be useful to read all of them before starting. Also think about the formulation and try to visualize the model. You are allowed to discuss the formulations, but have to make a note of the people you have discussed with. 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 documents on the web in general. Although you are allowed to discuss the problem formulations with others, you are not allowed to discuss solutions, and any discussions concerning the problem formulations must be described in the solutions you hand in. 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 scientific as well as corporate activities. 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!

Some questions in this assignment requires you to use the data we generated. You can access to the data via this link: `https://gits-15.sys.kth.se/koptagel/AdvML20`.

The grading of the assignment will be as follows,

- **E** Correctly completed **two** problems of the four 2.1, 2.2, 2.3, and 2.4.
- **D** Correctly completed **three or four** problems of the four 2.1, 2.2, 2.3, and 2.4.

Good Luck!

2.1 Dependencies in a Directed Graphical Model

Consider the graphs shown in the below figures. In the below question, consider independence as independence in all distributions that factorize according to the graph and dependence as dependence in some such distribution. You merely have to answer "yes" or "no" to each question.

Question 2.1.1: *In the graphical model of Figure 1, is* $\mu_{r,c} \perp \mu_{r,c+1}$*?*

Question 2.1.2: *In the graphical model of Figure 1, is* $X_{r,c} \perp X_{r,c+1} | \{\mu_{r,c}, \mu_{r,c+1}\}$*?*

Question 2.1.3: *Consider the graphical model of Figure 1, give a minimal set of variables* A *such that* $X_{r,c} \perp \mu_0 | A$.

Question 2.1.4: *Consider the graphical model of Figure 2, is* $Z \perp X | C$ *where* $Z = \{Z_m^n : n \in [N], m \in [M]\}$, $X = \{X_m^n : n \in [N], m \in [M]\}$, *and* $C = \{C^n : n \in [N]\}$*? (Notice that* $[N]$ *denotes* $\{1, \ldots, N\}$ *etcetera.)*

Question 2.1.5: *Consider the graphical model of Figure 2, is* $A \perp e | B$ *where* $A = \{A_{i,j}^k : k \in [K], i, j \in [I]\}$ *and* $e = \{e_{i,s}^k : k \in [K], i \in [I], r \in [R]\}$, *and* $B = \{Z_m^n : m \in [M], m$ *odd*$\} \cup \{X_m^n : m \in [M], m$ *even*$\}$.

Question 2.1.6: *Consider the graphical model of Figure 2, give a minimal set of variables* B *such that* $A \perp X | B$ *where* $A = \{A_{i,j}^k : k \in [K], i, j \in [I]\}$ *and* $X = \{X_m^n : n \in [N], m \in [M]\}$.

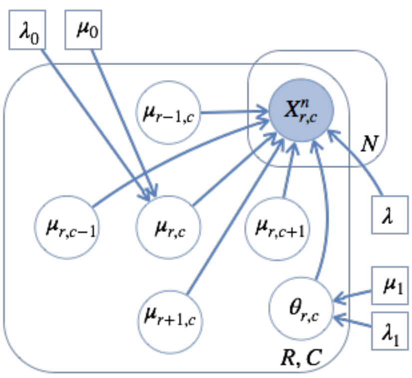

Figure 1: Graphical Model for the leaky units models. All incoming edges for variables with index r, c are shown. That is, these edges exist for all values of r and c.

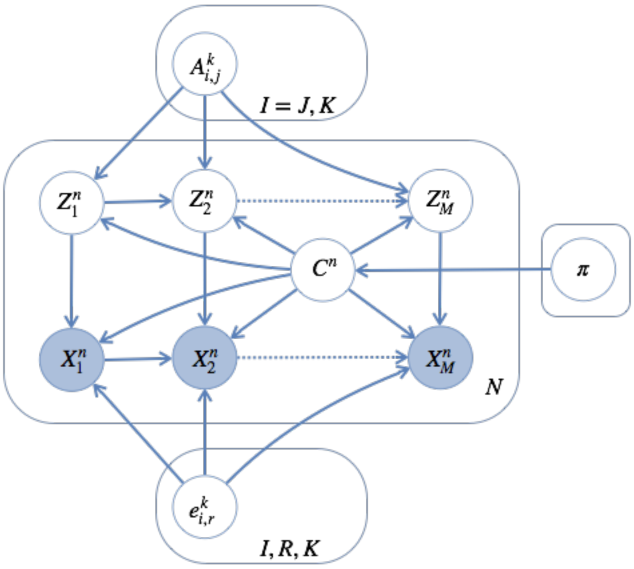

Figure 2: Mixture of HMMs.

2.2 Likelihood of a Tree Graphical Model

Let T be a binary tree, with vertex set $V(T)$ and leaf set $L(T)$, and consider the graphical model T, Θ described as follows. For each vertex $v \in V(T)$ there is an associated random variable X_v that assumes values in $[K]$. Moreover, for each $v \in V(T)$, the CPD $\theta_v = p(X_v|x_{\text{pa}(v)})$ is a categorical distribution. Let $\beta = \{x_l : l \in L(T)\}$ be an assignment of values to all the leaves of T.

Question 2.2.7: *Implement a dynamic programming algorithm that, for a given T, Θ and β, computes $p(\beta|T, \Theta)$.*

Question 2.2.8: *Apply your algorithm to the graphical model and data provided separately.*

2.3 Simple Variational Inference

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

Question 2.3.10: *What is the exact posterior?*

Question 2.3.11: *Compare the inferred variational distribution with the exact posterior. Run the inference on data points drawn from iid Gaussians. Do this for three interesting cases and visualize the results. Describe the differences.*

2.4 Mixture of trees with observable variables

Consider the mixture model $\mathcal{M} = (\pi, \tau)$, where π is a categorical distribution on $[K]$ and $\tau = \{(T_k, \Theta_k) : k \in [K]\}$ is a set of K graphical models that each is a tree with vertices V and root r. All CPDs are binary and all variables are observable. There is an EM algorithm that, for given data $\mathcal{D} = \{x^n : n \in [N]\}$, estimates $\mathcal{M}$ by iteratively performing the following steps w.r.t. to a current $\mathcal{M} = \pi, \tau$.

1. For each n, k, compute the responsibilities

$$r_{n,k} = \pi_k \, p(x^n | T_k, \Theta_k) \; / \; p(x^n).$$

2. Set $\pi'_k = \sum_{n=1}^{N} r_{n,k}/N$.

3. For each k, let G_k be a directed graph with edge weights defined by $w(st) = I_q(X_s, X_t)$, where $I_{q^k}(X_s, X_t)$ is the mutual information between X_s and X_t under the distribution q^k, i.e.,

$$I_{q^k}(X_s, X_t) = \sum_{a,b \in \{0,1\}} q^k(X_s = a, X_t = b) \log \frac{q^k(X_s = a, X_t = b)}{q^k(X_s = a) q^k(X_t = b)},$$

   and q^k is defined by

$$q^k(X_s = a, X_t = b) = \frac{\sum_{n \in [N] : X_s^n = a, X_t^n = b} r_{n,k}}{\sum_{n \in [N]} r_{n,k}}.$$

   Moreover, any term in $I_{q^k}(X_s, X_t)$ for which $q^k(X_s = a, X_t = b) = 0$ is considered to be 0.

4. Let T'_k be a maximum spanning tree in G_k.

5. Let $\Theta'_k(X_r) = q^k(X_r)$ and $\Theta'_k(X_s = a | X_t = b) = q^k(X_s = a | X_t = b)$.

The root stays the same; it is facilitating our computations, but any root would give the same result. Initialize the EM algorithm randomly, independently of the data, and use sieving. If you run into problems with values being zero, you can change them slightly using the Python "sys" module that contains a function called float_info. (e.g., *sys.float_info.epsilon: Value: $2e^{-16}$ or *sys.float_info.min: Value: $2e^{-308}$.)

Question 2.4.12: *Implement this EM algorithm.*

Question 2.4.13: *Apply your algorithm to the provided data and show how well you reconstruct the mixtures. First, compare the real and inferred trees with the* unweighted Robinson-Foulds *(aka* symmetric difference*) metric. Do the trees have similar structure (don't worry if the inferred trees don't match with the real trees)? Then, compare the likelihoods of real and inferred mixtures.*

Question 2.4.14: *Simulate new tree mixtures with different number of nodes, samples and clusters. Try to find some interesting cases. Analyse your results as in the previous question.*