DD2434/FDD3434 Machine Learning, Advanced Course
Assignment 2, 2021 (V2.0)

Jens Lagergren

Deadline, see Canvas

Read this before starting

You will present the assignment by a written report in PDF format, 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 (including whom you discussed with).

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 and some helper functions via this link:
https://gits-15.sys.kth.se/koptagel/AdvML21.

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 graphical models shown in Figures 1 and 2. You merely have to answer "yes" or "no" to each question.

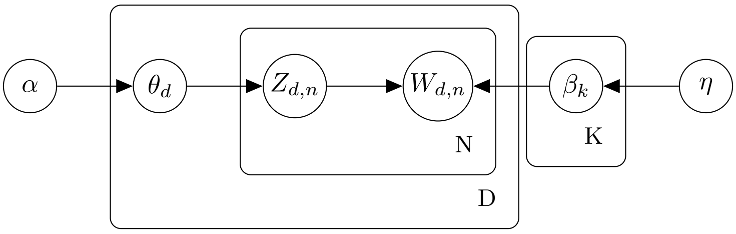

Figure 1: Graphical model of smoothed LDA.

Question 2.1.1: *In the graphical model of Figure 1, is* $W_{d,n} \perp W_{d,n+1} \mid \theta_d, \beta_{1:K}$*?*

Question 2.1.2: *In the graphical model of Figure 1, is* $\theta_d \perp \theta_{d+1} \mid Z_{d,1:N}$*?*

Question 2.1.3: *In the graphical model of Figure 1, is* $\theta_d \perp \theta_{d+1} \mid \alpha, Z_{1:D,1:N}$*?*

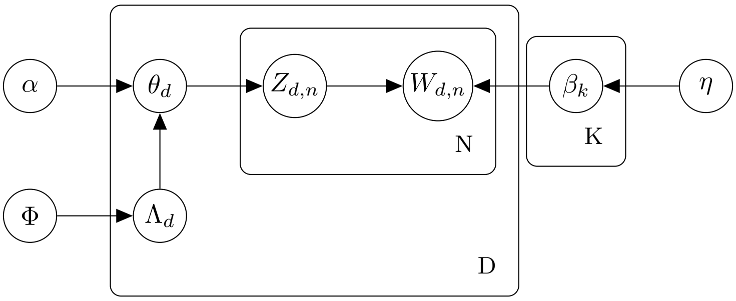

Figure 2: Graphical model of Labeled LDA.

Question 2.1.4: *In the graphical model of Figure 2, is* $W_{d,n} \perp W_{d,n+1} \mid \Lambda_d, \beta_{1:K}$*?*

Question 2.1.5: *In the graphical model of Figure 2, is* $\theta_d \perp \theta_{d+1} \mid Z_{d,1:N},\ Z_{d+1,1:N}$*?*

Question 2.1.6: *In the graphical model of Figure 2, is* $\Lambda_d \perp \Lambda_{d+1} \mid \Phi, Z_{1:D,1:N}$*?*

2.2 Likelihood of a Tree Graphical Model

Let T be a rooted 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: *Report $p(\beta | T, \Theta)$ for each given data, separately.*
You should report 15 values in total; 5 for small tree, 5 for medium tree and 5 for large tree.

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 i.i.d. Gaussians. Do this for three interesting cases and visualize the results. Describe the differences.*

2.4 Super Epicentra - Expectation-Maximization

We have seismographic from an area with frequent earthquakes emanating from K super epicentra. Each super epicentra is modeled by a 2-dimensional Gaussian determining the location of an earthquake and a Poisson distribution determining its strength. As shown in Figure 3 the entire model is a mixture of K such components. The variable Z^n is a class variable that follows a categorical distribution π and determines the super epicentra of the nth observation, i.e., the parameters of the Gaussian distribution that X^n is sampled from, $\mu_k = (\mu_{k,1}, \mu_{k,2})$ and $\tau_k = (\tau_{k,1}, \tau_{k,2})$ where $\tau_{k,1}$ and $\tau_{k,2}$ are diagonal elements of the diagonal precision matrix, as well as the intensity of the Poisson distribution that S^n is sampled from, λ_k .

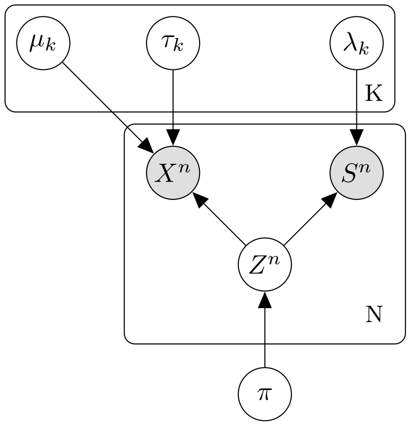

Figure 3: Mixture of components modeling location and strengths of earthquakes associated with a super-epicentra. In the figure, $\mu_k = (\mu_{k,1}, \mu_{k,2})$ and $\tau_k = (\tau_{k,1}, \tau_{k,2})$.

Question 2.4.12: *Derive an EM algorithm for the model.*

Question 2.4.13: *Implement your EM algorithm.*

Question 2.4.14: *Apply it to the data provided, give an account of the success, and provide visualizations for a couple of examples. Repeat it for a few different choices of K.*