DD2434/FDD3434 Machine Learning, Advanced Course
Assignment 1A, 2022

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. The assignment should be done in groups of two, 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 other groups, you are not allowed to discuss solutions, and any discussions concerning the problem formulations must be described in the solutions you hand in (including which group 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 an 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 1A and 2A will be as follows,

- **E** 30-44 points, where at least 20 points are from Assignment 1A and 10 points from 2A.
- **D** 45-60 points, where at least 20 points are from Assignment 1A and 10 points from 2A.
- All points over 30 will be counted as bonus points for assignment 1B and 2B.

Good Luck!

1.1 Exponential Family

A number of common distributions can be rewritten as exponential-family distributions with natural parameters, in the following form:

$$p(x|\boldsymbol{\theta}) = h(x) \exp\Big(\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \boldsymbol{T}(x) - A(\boldsymbol{\eta})\Big)$$

Below we provide five different distributions from exponential-family. Show which common distributions they correspond to. (1 point per correct answer)

Question 1.1.1:

- $\theta = \lambda$
- $\eta(\theta) = \log \theta$
- $h(x) = \frac{1}{x!}$
- $T(x) = x$
- $A(\eta) = e^{\eta}$

Question 1.1.2:

- $\boldsymbol{\theta} = [\alpha,\ \beta]$
- $\boldsymbol{\eta}(\boldsymbol{\theta}) = [\theta_1 - 1, -\theta_2]$
- $h(x) = 1$
- $\boldsymbol{T}(x) = [\log x,\ x]$
- $A(\boldsymbol{\eta}) = \log \Gamma(\eta_1 + 1) - (\eta_1 + 1) \log(-\eta_2)$

Question 1.1.3:

- $\boldsymbol{\theta} = [\mu,\ \sigma^2]$
- $\boldsymbol{\eta}(\boldsymbol{\theta}) = [\frac{\theta_1}{\theta_2}, -\frac{1}{2\theta_2}]$
- $h(x) = \frac{1}{\sqrt{2\pi}}$
- $\boldsymbol{T}(x) = [x,\ x^2]$
- $A(\boldsymbol{\eta}) = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2} \log(-2\eta_2)$

Question 1.1.4:

- $\theta = \lambda$
- $\eta(\theta) = -\theta$
- $h(x) = 2$
- $T(x) = x$
- $A(\eta) = -\log(-\eta/2)$

Question 1.1.5:

- $\boldsymbol{\theta} = [\psi_1, \psi_2]$
- $\boldsymbol{\eta}(\boldsymbol{\theta}) = [\theta_1 - 1,\ \theta_2 - 1]$
- $h(x) = 1$
- $\boldsymbol{T}(x) = [\log x,\ \log(1-x)]$
- $A(\boldsymbol{\eta}) = \log\Gamma(\eta_1 + 1) + \log\Gamma(\eta_2 + 1) - \log\Gamma(\eta_1 + \eta_2 + 2)$

1.2 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. Note that the words, $W_{d,n}$, are treated as observed only if the question explicitly states it. (1 point per correct answer)

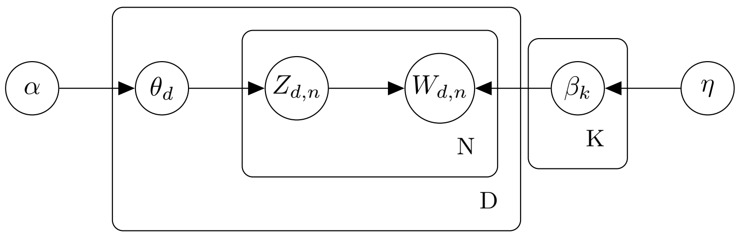


Figure 1: Graphical model of smoothed LDA.

Question 1.2.6: *In the graphical model of Figure 1, is* $W_{d,n} \perp W_{d,n+1} \mid \theta_d, \beta_{1:K}$*?*

Question 1.2.7: *In the graphical model of Figure 1, is* $\theta_d \perp \theta_{d+1} \mid Z_{d,1:N}$*?*

Question 1.2.8: *In the graphical model of Figure 1, is* $\theta_d \perp \theta_{d+1} \mid \alpha, Z_{1:D,1:N}$*?*

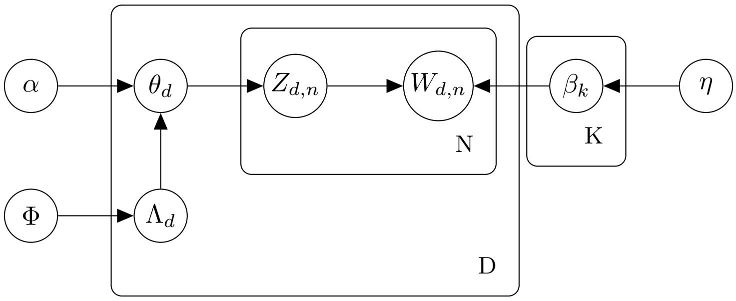


Figure 2: Graphical model of Labeled LDA.

Question 1.2.9: *In the graphical model of Figure 2, is* $W_{d,n} \perp W_{d,n+1} \mid \Lambda_d, \beta_{1:K}$*?*

Question 1.2.10: *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 1.2.11: *In the graphical model of Figure 2, is* $\Lambda_d \perp \Lambda_{d+1} \mid \Phi, Z_{1:D,1:N}$*?*

1.3 CAVI

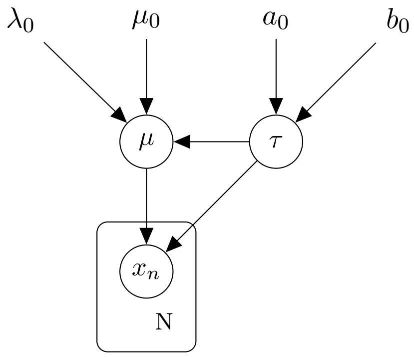


Figure 3: DGM

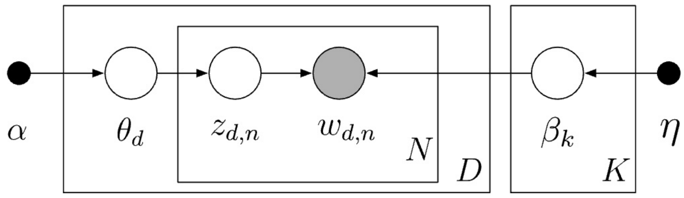
Consider the model defined by Equation (10.21)-(10-23) in Bishop, for which DGM is presented in Figure 3. We are here concerned with the VI algorithm for this model covered during the lectures and in the book.

Question 1.3.12: *Implement a function that generates data points for the given model. Set $\mu = 1$, $\tau = 0.5$ and generate datasets with size N=10,100,1000. Plot the histogram for each of 3 datasets you generated. (2 points)*

Question 1.3.13: *Find ML estimates of the variables μ and τ. (1 points)*

Question 1.3.14: *What is the exact posterior? (2 points)*

Question 1.3.15: *Implement the VI algorithm for the variational distribution in Equation (10.24) in Bishop. Run the VI algorithm on the datasets. Compare the inferred variational distribution with the exact posterior and the ML estimate. Visualize the results and discuss your findings. (10 points)*



Var	Type	Conditional	Param	Relevant Expectations
z_{dn}	Multinomial	$\log\theta_{dk} + \log\beta_{k,w_{dn}}$	ϕ_{dn}	$\mathbb{E}[Z^k_{dn}] = \phi^k_{dn}$
θ_d	Dirichlet	$\alpha + \sum_{n=1}^{N} z_{dn}$	γ_d	$\mathbb{E}[\log\theta_{dk}] = \Psi(\gamma_{dk}) - \sum_{j=1}^{K}\Psi(\gamma_{dj})$
β_k	Dirichlet	$\eta + \sum_{d=1}^{D}\sum_{n=1}^{N} z^k_{dn} w_{dn}$	λ_k	$\mathbb{E}[\log\beta_{kv}] = \Psi(\lambda_{kv}) - \sum_{y=1}^{V}\Psi(\lambda_{ky})$

Figure 4: LDA DGM and conditional distributions (taken from Hoffman et al. 2013)

1.4 SVI - LDA

This assignment concerns SVI as presented in the Hoffman paper in general, and in particular SVI for the LDA model shown in 2.

Question 1.4.16: *What is the definition of local hidden variables according to the Hoffman paper? Answer using conditional probability distributions and notation: x_n for observations, z_n for local hidden variables, β for global hidden variables and α for fixed parameters.* *(1 points)*

Question 1.4.17: *Which are the global and local hidden variables of the LDA model in 4?* *(1 points)*

Question 1.4.18: *Write the ELBO for the LDA model as a function of variational parameters and natural parameters of the full conditionals. No derivation is needed, only the final expression. You may use an online source for this derivation, in which case you must provide the link.* *(2 points)*

Question 1.4.19: *Adjust the CAVI updates provided in the notebook to SVI updates and implement the SVI algorithm. Apply it to the data provided, give an account of the success, and provide visualizations for the provided datasets. Also, present a plot showing the runtime of each experiment.* *(7 points)*

1.5 BBVI

In BBVI without Rao-Blackwellization and control variates, the gradient is estimated using Monte-Carlo sampling, the score function of q and the joint of p.

Question 1.5.20: *For the simple model, $X|\theta, \sigma^2 \sim \mathcal{N}(\theta, \sigma^2), \theta \sim Gamma(\alpha, \beta)$ and σ^2 fixed, derive the gradient estimate w.r.t. ν using one sample $z \sim q(\theta)$, $q(\theta) = LogNormal(\nu, \epsilon^2)$ (2 points)*

Question 1.5.21: *Describe in one sentence what the Control variates are used for in the BBVI paper. (1 points)*