DD2434/FDD3434 Machine Learning, Advanced Course
Assignment 2A, 2024

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, 2A and 3A will be as follows,

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

Good Luck!

2.1 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 four different distributions from exponential-family. Show which common distributions they correspond to. (1 point per correct answer)

Question 2.1.1:

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

Question 2.1.2:

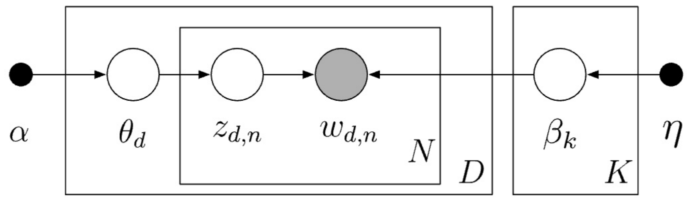
- $\boldsymbol{\theta} = [\alpha,\ \beta]$
- $\boldsymbol{\eta}(\boldsymbol{\theta}) = [\theta_1 - 1, -\theta_2]$
- $h(x) = 1$
- $\boldsymbol{T}(x) = [\log x,\ x]$
- $A(\boldsymbol{\eta}) = \log\Gamma(\eta_1 + 1) - (\eta_1 + 1)\log(-\eta_2)$

Question 2.1.3:

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

Question 2.1.4:

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



Var	Type	Conditional	Param	Relevant Expectations
z_{dn}	Multinomial	$\log \theta_{dk} + \log \beta_{k,w_{dn}}$	ϕ_{dn}	$\mathbb{E}[Z_{dn}^k] = \phi_{dn}^k$
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	$\eta + \sum_{d=1}^{D} \sum_{n=1}^{N} z_{dn}^k w_{dn}$	λ_k	$\mathbb{E}[\log \beta_{kv}] = \Psi(\lambda_{kv}) - \sum_{y=1}^{V} \Psi(\lambda_{ky})$

Figure 1: LDA DGM and conditional distributions (taken from Hoffman et al. 2013)

2.2 SVI - LDA

This assignment concerns SVI as presented in the Hoffman paper in general, and in particular SVI for the LDA model shown in 1.

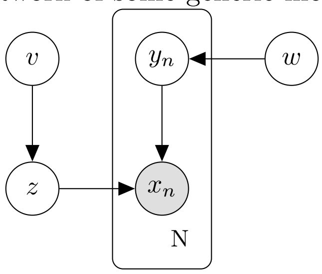
Question 2.2.5: *What is the definition of local hidden variables according to the Hoffman paper? Answer using conditional probability distributions and notation: x_n for observations, z_n for local hidden variables, β for global hidden variables and α for fixed parameters. (1 points)*

Question 2.2.6: *Consider the LDA model in 1. Let $w_d = w_{d,1:N}$ and $z_d = z_{d,1:N}$. Show that θ_d, z_d fulfills the definition in 2.2.5. (1 points)*

Question 2.2.7: *Write the ELBO for the LDA model as a function of variational parameters, $\phi_{dn}, \gamma_d, \lambda_k$, prior parameters, α, η and hyperparameters D, N, K, W. No derivation is needed, only the final expression. You may use an online source for this derivation, in which case you must provide a link to the source. (2 points)*

Question 2.2.8: *Adjust the CAVI updates provided in the notebook to SVI updates and implement the SVI algorithm. Use the function provided for generating data and run the algorithm for the cases defined in the notebook. In one sentence, comment the success and runtime of each experiment. (7 points)*

Figure 2: Bayesian network of some generic model in question 2.3.13.



2.3 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 2.3.9: *Let $X = (X_1, ..., X_N)$ be i.i.d. with $X_n|\theta, \sigma^2 \sim \mathcal{N}(\theta, \sigma^2), \theta \sim LogNormal(\nu, \epsilon^2)$ and σ^2 fixed. Write an expression for the Naive BBVI gradient estimate w.r.t. α using one sample $z \sim q(\theta)$, $q(\theta) = Gamma(\alpha, \beta)$ (2 points)*

Question 2.3.10: *Describe in one sentence what Rao-Blackwellization is used for in the BBVI paper. (1 points)*

Question 2.3.11: *Given the model in figure 2, and the mean-field appproximation: $q(y, w, z, v) = q_{\lambda^1}(w)q_{\lambda^2}(z)q_{\lambda^3}(v)\prod_n q_{\lambda_n^4}(y_n)$, describe qualitatively how the Rao-Blackwellized partial gradient of the ELBO w.r.t. λ_n^4, $\nabla_{\lambda_n^4}\mathcal{L}$ is obtained. Write out the final expression for the Rao-Blackwellized $\nabla_{\lambda_n^4}\mathcal{L}$. (2 points)*