DD2434/FDD3434 Machine Learning, Advanced Course
Assignment 1B, 2024

Jens Lagergren

Deadline, see Canvas

Read this before starting

There are some commonalities between the problems and they cover different aspects of the course and vary in difficulty, 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.

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 problems below can give 60 points in total. The grade thresholds of assignments 1B and 2B are given below. Note that you can have 30 bonus points from assignment 1A, 2A and 3A and 30 points from 2B.

- **D** 30 points.
- **C** 50 points.
- **B** 70 points.
- **A** 90 points.

These grades are valid for assignments submitted before the deadline, late assignments can at most receive the grade E, which makes it meaningless to hand in late solutions for this assignment.

Good Luck!

1.1 Dice model - CAVI

Two dice, π^1 and π^2, of dimensions D^1 and D^2, are generated from two Dirichlet distributions with parameters α^1 and α^2, respectively.

Furthermore, K dice, $\rho_1, ..., \rho_K$, all of dimension D^3, are generated independently from K Dirichlet distributions with parameters $\beta_1, ..., \beta_K$, respectively.

For each data point, X_n, where $n \in 1, ..., N$, two 'class' variables Z_n^1 and Z_n^2 are drawn from π^1 and π^2, respectively. Then, J number of X_{nj} are drawn i.i.d. from the dice ρ_k, where k is determined by $Z_n^1 + Z_n^2$.

The model distributions are summarized below:

- $p(\pi^1) = Dirichlet(\alpha^1)$
- $p(\pi^2) = Dirichlet(\alpha^2)$
- $p(Z_n^1|\pi^1) = Categorical(\pi^1)$
- $p(Z_n^2|\pi^2) = Categorical(\pi^2)$
- $p(X_n|Z_n^1 + Z_n^2 = k, \rho_k) = Categorical(\rho_k)$
- $p(\rho_k) = Dirichlet(\beta_k)$

Question 1.1.1: *Draw the DAG/PGM/Bayesian Network of the model described above.* *(X points)*

Question 1.1.2: *Express* $\log p(X, Z^1, Z^2, \pi^1, \pi^2, \rho|\alpha^1, \alpha^2, \beta)$ *in terms of indicator functions and sums of the log of known pdfs/pmfs.*

Do not insert the expressions for the pdfs, e.g., write "... + $\log p(\pi^1)$*", not the log of the actual expression of the Dirichlet pdf.* *(X points)*

Question 1.1.3: *Use the mean-field assumption:*

$$q(Z^1, Z^2, \pi^1, \pi^2, \rho) = q(\pi^1)q(\pi^2)\prod_n q(Z_n^1)q(Z_n^2)\prod_k q(\rho_k)$$

and the CAVI update equation to derive the optimal variational distributions and parameters of $q(Z_n^1)$, $q(\pi^1)$ *and* $q(\beta_k)$ *(X points)*

.

1.2 VAE for latent representation and image generation

Notebook "HT24-1B-VAE.ipynb" contains a partially completed implementation of a VAE to the MNIST dataset. Answer the questions 5.1 and 5.2 by reporting your derivations, the questions 9.1 and 10.4 by your discussion and argumentation, and the other questions by inserting your code into the relevant functions. Follow the guidelines in the notebook to train the model and generate the results. (X points)

1.3 Reparameterization of common distributions

Even though Gaussian distributions are easy to reparameterize and convenient to work with, they are not always the best fit for your model. For instance, your model may require the latent variable to be non-negative, or in $[0, 1]$. In order to apply VAE for those cases, you should know how to reparameterize some other common distributions. In this question you will apply the reparameterization trick to two different distributions (see the notebook "1B-Reparameterization.ipynb"):

Beta Distribution

> **Question 1.3.4:** *It is not straight-forward to reparameterize the Beta distribution. Therefore, it is common to approximate it by the Kumaraswamy distribution:*
> $Beta(a, b) \approx Kumaraswamy(a, b)$
> *Show that you can reparameterize* $Kumaraswamy(a, b)$ *(Hint: use inverse transform method).*
> *Implement your reparameterization method in Q1 of the notebook. (X points)*

Dirichlet Distribution

> **Question 1.3.5:** *Describe how to reparameterize the Dirichlet distribution by approximating it with softmax Gaussian distribution. You may use the paper Autoencoding Variational Inference For Topic Models as a reference (see section 3.2).*
> *Implement your reparameterization method in Q2 of the notebook. (X points)*