DD2434/FDD3434 Machine Learning, Advanced Course
Assignment 1B, 2023

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 1.1-1.4 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 and 2A 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 CAVI for Earth quakes

Solve either alternative 1 or alternative 2 below. Alternative 2 is simpler as it doesn't involve 2D vectors. Solving alternative 1 may yield 4 additional points for the CAVI updates.

Alternative 1. (in two dimensions) In an area with frequent earthquakes emanating from K super epicentra, we gather seismographic data on the strength, S_n, and the 2D-coordinates, X_n, of each outbreak. We introduce a class variable, Z_n, with a categorical distribution parameterized by π, which assigns the nth observation to a super epicentra. We model $S_n|Z_n = k, \lambda_k$ as a Poisson random variable with super epicentra specific intensity, λ_k, and $X_n|Z_n = k, \mu_k, \tau$ as a 2D Normal r.v. with super epicenter specific mean vector $\mu_k = (\mu_0, \mu_1)$ and precision matrix set to $\tau \cdot I$, where I is the 2D identity matrix. We set a 2D Normal prior on μ_k with mean vector $\nu = (\nu_0, \nu_1)$ and precision matrix $\rho \cdot I$. We set a Gamma prior on λ_k with shape parameter α and rate parameter β. The remaining parameters are treated as constants.

Alternative 2. (in one dimension) In an area with frequent earthquakes emanating from K super epicentra, we gather seismographic data on the strength, S_n, of each outbreak and measure the approximate radial distance from an antenna to each epicenter, R_n. The angular coordinate is not measurable by the antenna. S_n is modeled by a Poisson distribution with super epicentra specific intensity, λ_k. R_n is modeled by a Log-normal distribution with a super epicentra specific distance μ_k and fixed precision τ determined by the antenna. Furthermore, we introduce the class variable Z_n with a categorical distribution parameterized by π, which determines the super epicentra of the nth observation and the intensity of the Poisson distribution. That is, if $Z_n = k$, then the mean parameter of the Log-normal distribution that R_n is sampled from is μ_k and the intensity of the Poisson distribution that S^n is sampled from is λ_k. We set a 1D Normal prior on μ_k with mean ν and precision ρ. We set a Gamma prior on λ_k with shape parameter α and rate parameter β. The remaining parameters are treated as constants.

Question 1.1.1: *Draw the DGM of the model described above. (2 points)*

Question 1.1.2: *Express* $\log p(X, S, Z, \lambda, \mu|\pi, \tau, \alpha, \beta, \nu, \rho)$ *or* $\log p(R, S, Z, \lambda, \mu|\pi, \tau, \alpha, \beta, \nu, \rho)$ *in terms of known pdfs/pmfs. (2 points)*

Question 1.1.3: *Use the mean-field assumption* $q(Z, \mu, \lambda) = \prod_n q(Z_n) \prod_k q(\mu_k)q(\lambda_k)$ *and CAVI update equation and derive the optimal variational distributions and parameters of* $q(Z_n)$, $q(\mu_k)$ *and* $q(\lambda_k)$ *(16 points)*

.

1.2 VAE image generation

Notebook "1B-VAE.ipynb" contains a partially completed implementation of a VAE for the MNIST dataset. Answer the questions 5.1 and 5.2 by reporting your derivations and the other questions by inserting your code into the relevant functions. Follow the guidelines in the notebook to train the model and generate images. (15 points)

1.3 Reparametrization and the score function

In modules 5 and 6 you were introduced to two estimators of the gradient of the ELBO: one high variance estimator which uses the score function (the REINFORCE estimator) and one lower variance estimator which uses the reparameterization trick (the reparameterized gradient estimator). In the paper Sticking the Landing, Roeder et al. show that the reparametrized gradient estimator also contains a score function, which effects the variance of the estimator. *Hint: For this assignment you will never have to read beyond page 3 of the paper in question.*

Question 1.3.4: *Following the Sticking the Landing paper, that is you may use the equations of the paper, decompose the gradient of the ELBO using the reparameterization trick to a form where the score function appears as a term. (3 points)*

Question 1.3.5: *Show that the expectation of the score function is zero. (4 points)*

Question 1.3.6: *What solution do the authors propose to handle this score function? Answer in one sentence. (2 points)*

Question 1.3.7: *The authors mention that for particular cases, the score function may actually decrease the variance. What concept, introduced in our course and mentioned in the paper, describes how the score function acts in this situation? State the answer without explanation. (1 points)*

1.4 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"):

Exponential Distribution

Question 1.4.8: *Show how to reparameterize the exponential distribution with the parameter λ. Implement your reparameterization method in Q1 of the notebook. (6 points)*

Categorical Distribution

Question 1.4.9: *Describe how to reparameterize the categorical distribution by, (1.) Approximating it by the Gumbel-Softmax distribution, which then allows for differentation during training, and (2.) using the argmax function for evaluation. You may use the paper Categorical Reparameterization with Gumbel-Softmax as a reference.*
Implement your reparameterization method in Q2 of the notebook. (9 points)