# DD2434/FDD3434 Machine Learning, Advanced Course Module 3 Exercise

## November 2023

## Contents

4	Sto	chastic Variational Inference – Exercises	2
	4.1	Exponential family	2
	4.2	Metrics for distributions	2
	4.3	Mixture Model with Bernoulli observations - SVI	2
	4.4	Gaussian mixture model	:

Instead of presenting a summary of the subject, these exercise problems will refer to the Hoffman paper.

#### Stochastic Variational Inference – Exercises 4

#### 4.1 Exponential family

Show that the following distributions are in the exponential family:

- a) Beta-distribution
- b) Binomial-distribution
- c) Dirichlet-distribution
- d) Multinomial-distribution
- e) LogNormal-distribution

#### 4.2 Metrics for distributions

Let  $p_1 = p(X|\mu_1, \sigma_1^2) = Normal(\mu_1, \sigma_1^2)$  and  $p_2 = p(Y|\mu_1, \sigma_1^2) = Normal(\mu_2, \sigma_2^2)$ . Write a function that calculates the euclidean distance, d, between the pdf's and another function that calculates symmetric KL-divergence,  $D_{KL}^{sym}$ , between the pdf's.

Visualize  $p_1$  and  $p_2$  for the cases:

- 1)  $\mu_1 = 0, \mu_2 = 0.1, \sigma_1^2 = \sigma_2^2 = 0.01$ 2)  $\mu_1 = 0, \mu_2 = 10, \sigma_1^2 = \sigma_2^2 = 10^4$

Calculate and compare  $d(p_1, p_2)$  and  $D_{KL}^{sym}(p_1, p_2)$  between the scenarios. Which metric do you think best distinguishes  $p_1$  and  $p_2$ ?

#### 4.3 Mixture Model with Bernoulli observations - SVI

Consider the Bernoulli Mixture model (MM) of module 3.

- a) Which are the local and global latent variables of the model?
- b) Derive the full conditionals of the model.
- c) Show that these distributions are in the exponential family.
- d) Assume the mean-field factorization:  $q(Z,\pi,\theta) = q(\pi) \prod_n q(Z_n) \prod_k q(\theta_k)$  and that each variational distribution is in the same parametric family as the associated complete conditional. Use the natural parameters from c) and equations 15 and 16 of Hoffman to derive the updates of the global and local latent variables. How do these updates compare to that of the Bernoulli-MM of Module 3?
- e) Use the code from the Bernoulli-MM of exercises of module 3 and rewrite the CAVI algorithm to the SVI algorithm described in Figure 4 of Hoffman. Compare the outputs of the SVI algorithm and CAVI algorithm for different parameters. How do they compare? Compare the run time for a fixed number of iterations, large N  $(N \ge 10^6)$  and large D  $(D \ge 10^3)$ .

### 4.4 Gaussian mixture model

A Gaussian mixture model with K components can be written:

$$p(x|\mu,\sigma^2) = \sum_k \pi_k \mathcal{N}(\mu_k, \sigma_k^2)$$
 (1)

where  $\pi_k$  are the mixture weights. Note that this is the formulation of the GMM without latent component assignment variables,  $Z_n$ .

Try (and fail) rewriting  $p(x|\mu, \sigma^2)$  in the exponential family form (it's not possible), e.g. by testing the case of K = 2. Briefly discuss where it goes wrong.