MA060129, Bayesian Machine Learning
Homework 1: Theoretical Problems

{Evgenii.Egorov, A.Kuzina}@skoltech.ru

September 2019

Content:

- Problem 1 [20 pts]
- Problem 2 [20 pts]
- Problem 3 [20 pts]
- Problem 4 [20 pts]
- Problem 5 [20 pts]

Problem 1

Recall definition of exponential family distribution:

$$p(x;\lambda) = \exp\left(\langle t(x), \lambda\rangle - F(\lambda) + k(x)\right)$$
$$F(\lambda) = \log \int\limits_X \exp\left(\langle t(x), \lambda\rangle + k(x)\right) dx$$

Consider two distributions from the same class of exponential family:

$$p_1(x;\lambda_1) = \exp\left(\langle t(x), \lambda_1\rangle - F(\lambda_1)\right)$$
$$p_2(x;\lambda_2) = \exp\left(\langle t(x), \lambda_2\rangle - F(\lambda_2)\right)$$

For $\alpha \in [0;1]$ consider geometrical average distribution:

$$q(x;\alpha) = \frac{p_1^{\alpha}(x;\lambda_1)p_2^{1-\alpha}(x;\lambda_2)}{\int\limits_X p_1^{\alpha}(x;\lambda_1)p_2^{1-\alpha}(x;\lambda_2)dx}$$

Solve following problems.

Please note, that the answer should not contain expectation, integration or $t(x)$

1. Find entropy of p_1

$$H(p_1) = -\int\limits_X p_1(x;\lambda_1)\log p_1(x;\lambda_1)dx$$

2. Find KL-divergence between $p_1(x;\lambda_1)$ and $p_2(x;\lambda_2)$

$$KL(p_1||p_2) = \int\limits_X p_1(x;\lambda_2)\log\frac{p_1(x;\lambda_1)}{p_2(x;\lambda_2)}dx$$

3. Find KL-divergence between $p_1(x;\lambda_1)$ and $q(x;\alpha)$

$$KL(p_1||q) = \int\limits_X p_1(x;\lambda_1)\log\frac{p_1(x;\lambda_1)}{q(x;\alpha)}dx$$

Problem 2

Assume you have model parameterized by α with the uniform prior distribution.

Derive the prior distribution for α^2

Problem 3

Consider model with the following likelihood function:

$$p(X_1 = x_1, \dots X_K = x_K|\pi) = \frac{N!}{x_1!x_2!\dots x_K!}\pi_1^{x_1}\pi_2^{x_2}\dots\pi_K^{x_K}, \quad \sum_{k=1}^{K} x_k = N, \quad \sum_{k=1}^{K}\pi_k = 1$$

Let us choose the following prior for the model parameters π:

$$p(\pi_1, \dots \pi_k|\alpha) = \frac{\Gamma\left(\sum_{i=1}^{K}\alpha_i\right)}{\prod_{i=1}^{K}\Gamma(\alpha_i)}\prod_{i=1}^{K}\pi_i^{\alpha_i - 1}$$

Given i.i.d. observations $X = (x_i, \dots x_N)$, calculate:

1. MLE estimate of the model parameters
2. Posterior distribution and its expectation
3. MAP estimate
4. Predictive distribution
5. Re-do points (2-4) with the following prior:

   $p(\pi_1, \dots \pi_k|\alpha^{(1)}, \alpha^{(2)}, \gamma) = \gamma p_1(\pi_1, \dots \pi_k|\alpha^{(1)}) + (1-\gamma)p_2(\pi_1, \dots \pi_k|\alpha^{(2)})$,
   where $p_i(\pi_1, \dots \pi_k|\alpha^{(i)}) \sim \mathrm{Dir}(\alpha^{(i)})$

Problem 4

Consider the f-divergence between distributions P and Q with corresponding densities p, q

$$D_f(p||q) = \langle f\left(\frac{p}{q}\right)\rangle_q, \quad \text{where } f \text{ is convex function}$$

Consider the particular case with $f^*(x) = x\log x$. Show that $D_{f^*}(p||q)$ is convex over both p, q.

Problem 5

Consider the family of the distribution:

$$p_i(x) = \tfrac{1}{Z(\beta_i)}\pi(x)\exp(-\beta_i h(x)), \beta \in [\beta_n; \beta_0]$$

Prove the identity:

$$\log Z(\beta_n) - \log Z(\beta_0) = \int_{\beta_n}^{\beta_0}\langle h(x)\rangle_{p_\beta} d\beta$$