

## Homework 5

1. На лекции мы с вами изучили планарные потоки вида:

$$\mathbf{x} = g(\mathbf{z}, \boldsymbol{\theta}) = \mathbf{z} + \mathbf{u}h(\mathbf{w}^T \mathbf{z} + b), \quad (1)$$

где  $\mathbf{u} \in \mathbb{R}^m$ ,  $\mathbf{w} \in \mathbb{R}^m$ ,  $b \in \mathbb{R}$ . Существует естественное обобщение планарных потоков вида:

$$\mathbf{x} = g(\mathbf{z}, \boldsymbol{\theta}) = \mathbf{z} + \mathbf{V}h(\mathbf{W}^T \mathbf{z} + \mathbf{b}), \quad (2)$$

где  $\mathbf{V} \in \mathbb{R}^{m \times k}$ ,  $\mathbf{W} \in \mathbb{R}^{m \times k}$ ,  $\mathbf{b} \in \mathbb{R}^k$ . Такой поток получил название *Sylvester flow*.

(a) **(1 pt)** Докажите упрощенный вариант matrix-determinant lemma:

$$\det(\mathbf{I}_m + \mathbf{V}\mathbf{W}^T) = \det(\mathbf{I}_k + \mathbf{W}^T \mathbf{V}).$$

(b) **(1.5 pt)** Посчитайте детерминант матрицы Якоби для преобразования (2) и примените к нему доказанную в предыдущем пункте лемму.

(c) **(1 pt)** Для того чтобы уменьшить сложность подсчета полученного детерминанта авторы предложили применить QR-разложение вида:

$$\mathbf{V} = \mathbf{Q}\mathbf{U}; \quad \mathbf{W} = \mathbf{Q}\mathbf{L},$$

где  $\mathbf{Q} \in \mathbb{R}^{m \times k}$  – ортогональная матрица ( $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$ ),  $\mathbf{U} \in \mathbb{R}^{k \times k}$  – верхнетреугольная матрица,  $\mathbf{L} \in \mathbb{R}^{k \times k}$  – нижнетреугольная матрица. Выпишите выражение для детерминанта якобиана, используя это разложение.

(d) **(1.5 pt)** Вычислите и сравните сложности для подсчета детерминанта матрицы Якоби до применения леммы, после применения леммы и после применения QR-разложения

2. **(2 pt)** На лекции мы с вами доказали теорему об *операции над ELBO*:

$$\frac{1}{n} \sum_{i=1}^n KL(q(\mathbf{z}|\mathbf{x}_i)||p(\mathbf{z})) = KL(q(\mathbf{z})||p(\mathbf{z})) + \mathbb{I}_q[\mathbf{x}, \mathbf{z}],$$

где первый член  $KL(q(\mathbf{z})||p(\mathbf{z}))$  включает в себя агрегированное апостериорное распределение  $q(\mathbf{z})$  и априорное распределение  $p(\mathbf{z})$ . Наша цель сейчас – разобраться со вторым членом. На лекции второй член был равен:

$$\mathbb{I}_q[\mathbf{x}, \mathbf{z}] = \frac{1}{n} \sum_{i=1}^n KL(q(\mathbf{z}|\mathbf{x}_i)||q(\mathbf{z})). \quad (3)$$

На самом деле это совместная информация между  $\mathbf{x}$  и  $\mathbf{z}$  по эмпирическому распределению данных и распределению  $q(\mathbf{z}|\mathbf{x})$ . Представим индекс объекта выборки  $i$  - случайной величиной.

$$q(i, \mathbf{z}) = q(i)q(\mathbf{z}|i); \quad p(i, \mathbf{z}) = p(i)p(\mathbf{z}); \quad q(i) = p(i) = \frac{1}{n}.$$

$$q(\mathbf{z}|i) = q(\mathbf{z}|\mathbf{x}_i) \quad q(\mathbf{z}) = \sum_{i=1}^n q(i, \mathbf{z}) = \frac{1}{n} \sum_{i=1}^n q(\mathbf{z}|\mathbf{x}_i);$$

Совместная информация - это некая мера независимости двух случайных величин:

$$\mathbb{I}_q[\mathbf{x}, \mathbf{z}] = \mathbb{E}_{q(i, \mathbf{z})} \log \frac{q(i, \mathbf{z})}{q(i)q(\mathbf{z})}. \quad (4)$$

Докажите, что выражение (4) равно выражению (3).

3. **(2 pt)** Вспомним вариационную нижнюю оценку для задачи деквантизации дискретных данных:

$$\log P(\mathbf{x}|\boldsymbol{\theta}) \geq \int q(\mathbf{u}|\mathbf{x}) \log \frac{p(\mathbf{x} + \mathbf{u}|\boldsymbol{\theta})}{q(\mathbf{u}|\mathbf{x})} d\mathbf{u} = \mathcal{L}(q, \boldsymbol{\theta}).$$

Мы с вами обсуждали, что вариационную нижнюю оценку можно улучшить с помощью Importance Sampling. Выпишите нижнюю оценку для деквантизации, используя  $\mathcal{L}_k$  по аналогии с моделью IWAE (необходимо использовать не одну  $\mathbf{u}$ , а набор из  $\{\mathbf{u}_k\}_{k=1}^K$ ).

4. Стандартный GAN часто страдает от проблем с затухающим градиентом. Least Squares GAN пытается решить данную проблему заменой функции ошибки на следующую:

$$\begin{aligned} \min_D V(D) &= \min_D \frac{1}{2} [\mathbb{E}_{\pi(\mathbf{x})} (D(\mathbf{x}) - b)^2 + \mathbb{E}_{p(\mathbf{z})} (D(G(\mathbf{z})) - a)^2] \\ \min_G V(G) &= \min_G \frac{1}{2} [\mathbb{E}_{\pi(\mathbf{x})} (D(\mathbf{x}) - c)^2 + \mathbb{E}_{p(\mathbf{z})} (D(G(\mathbf{z})) - c)^2], \end{aligned}$$

где  $a, b, c \in \mathbb{R}$  – фиксированные константы.

- (a) **(1.5 pt)** Найдите формулу для оптимального дискриминатора  $D^*$ .  
 (b) **(1 pt)** Выпишите выражение функции ошибок генератора  $V(G)$  в случае оптимального дискриминатора  $D^*$ .  
 (c) **(1.5 pt)** Докажите, что при  $b - c = 1$ ,  $b - a = 2$ , функция ошибок генератора  $V(G)$  в случае оптимального дискриминатора  $D^*$  принимает вид:

$$V(G) = \frac{1}{2} \chi_{\text{Pearson}}^2(\pi(\mathbf{x}) + p(\mathbf{x}|\boldsymbol{\theta}) || 2p(\mathbf{x}|\boldsymbol{\theta})),$$

где  $\chi_{\text{Pearson}}^2(p||q)$  – квадратичная дивергенция Пирсона:

$$\chi_{\text{Pearson}}^2(p||q) = \int \frac{(p(\mathbf{x}) - q(\mathbf{x}))^2}{p(\mathbf{x})} d\mathbf{x}.$$