Машинное обучение, ФКН ВШЭ
Теоретическое домашнее задание №13

Задача 1. Для двух одномерных нормальных распределений $\mathcal{N}(x \mid \mu_1, \sigma_1)$, $\mathcal{N}(x \mid \mu_2, \sigma_2)$ найдите дивергенцию Кульбака-Лейблера.

$$\mathrm{KL}(\mathcal{N}(x \mid \mu_1, \sigma_1) \| \mathcal{N}(x \mid \mu_2, \sigma_2)) - ? \tag{0.1}$$

Задача 2. Рассмотрим общую схему EM-алгоритма, выводимую через разложение

$$\log p(X \mid \Theta) = \mathcal{L}(q, \Theta) + \mathrm{KL}\left(q \,\|\, p\right).$$

На E-шаге ищется распределение q, доставляющее максимум нижней оценке $\mathcal{L}(q, \Theta^{\mathrm{old}})$ при фиксированном Θ^{old}.

Модифицируем E-шаг: будем теперь искать максимум не среди всех возможных распределений, а лишь среди вырожденных, то есть присваивающих единичную вероятность одной точке и нулевую вероятность всем остальным. Как будут выглядеть E- и M-шаги в этом случае?

Задача 3. Наблюдается выборка бинарных значений $y = (y_1, \dots, y_n)$, $y_i \in \{0, 1\}$. Все элементы выборки генерируются независимо, но известно, что в некоторый момент z меняется частота генерации единиц. Т.е., для всех $i < z$ $P(y_i = 1) = \theta_1$, а для всех $i \geqslant z$ $P(y_i = 1) = \theta_2$. Необходимо вывести формулы для EM-алгоритма, где z–скрытая переменная, а θ_1, θ_2 – параметры распределений.