Машинное обучение.
Домашнее задание №1

Задача 1. Рассмотрим общую схему EM-алгоритма, выводимую через разложение

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

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 2. Рассмотрим пример, показывающий, что неполное правдоподобие для смеси распределений может иметь особые точки.

Возьмем одномерную выборку, состоящую из первых десяти натуральных точек: $X = \{1, 2, 3, \ldots, 10\}$. Будем настраивать на данную выборку смесь из двух нормальных распределений:

$$p(x) = \pi_1 \mathcal{N}(x \mid \mu_1, \sigma_1^2) + \pi_2 \mathcal{N}(x \mid \mu_2, \sigma_2^2).$$

Первую гауссиану выберем произвольным образом и зафиксируем. Центр второй гауссианы разместим в одном из объектов обучающей выборки, а дисперсию устремим к нулю: $\pi_1 = \pi_2 = 0.5$, $\mu_1 = 5.5$, $\sigma_1^2 = 1$, $\mu_2 = 1$, $\sigma_2^2 \to 0$. Покажите, что в этом случае правдоподобие будет стремиться к бесконечности.

Задача 3. Рассмотрим смесь распределений Бернулли:

$$p(x \mid \mu, \pi) = \sum_{k=1}^{K} \pi_k p(x \mid \mu_k),$$

где $x \in \mathbb{R}^d$, $\mu = \{\mu_1, \ldots, \mu_K\}$, $\mu_k \in [0, 1]^d$, $\pi = \{\pi_1, \ldots, \pi_K\}$, $\sum_{i=1}^{K} \pi_k = 1$, и

$$p(x_i \mid \mu_k) = \mu_{ki}^{x_i} (1 - \mu_{ki})^{1 - x_i},$$

$$p(x \mid \mu_k) = \prod_{i=1}^{d} \mu_{ki}^{x_i} (1 - \mu_{ki})^{1 - x_i}.$$

Иными словами, k-я компонента смеси — это такое распределение на d-мерных бинарных векторах, что i-я координата вектора имеет распределение Бернулли с параметром μ_{ki}.

Пусть дана выборка $X^\ell = \{x_1, \ldots, x_\ell\}$. Введите скрытые переменные по аналогии со смесями нормальных распределений, и выведите формулы для E- и M-шагов EM-алгоритма.

Задача 4. Рассмотрим смесь двух одномерных нормальных распределений:

$$p(x) = \pi_1 \mathcal{N}(x \mid \mu_1, \sigma_1^2) + \pi_2 \mathcal{N}(x \mid \mu_2, \sigma_2^2).$$

Дисперсии σ_1^2 и σ_2^2 известны.

Пусть дана выборка $X^\ell = \{x_1, \ldots, x_\ell\}$. Введите в модель скрытые переменные и выведите формулы шагов EM-алгоритма для настройки параметров $\pi_1, \pi_2, \mu_1, \mu_2$.

Задача 5. Пусть $p(x, y)$ и $q(x, y)$ — такие распределения, что

$$p(x, y) = p_1(x) p_2(y);$$
$$q(x, y) = q_1(x) q_2(y).$$

Докажите, что

$$\mathrm{KL}\,(q \parallel p) = \mathrm{KL}\,(q_1 \parallel p_1) + \mathrm{KL}\,(q_2 \parallel p_2)\,.$$

Задача 6. Рассмотрим два d-мерных нормальных распределения с единичными ковариационными матрицами:

$$p(x) = \mathcal{N}(x \mid 0, I);$$
$$q(x) = \mathcal{N}(x \mid \mu, I).$$

Докажите, что

$$\mathrm{KL}\,(q \parallel p) = \frac{\|\mu\|^2}{2}.$$