

Машинное обучение, весна

Домашнее задание №2

Задача 1. Рассмотрим пример, показывающий, что неполное правдоподобие для смеси распределений может иметь особые точки.

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

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

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

Задача 2. Рассмотрим смесь распределений Бернулли:

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

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

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

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 $X^\ell = \{x_1, \dots, x_\ell\}$. Введите скрытые переменные по аналогии со смесями нормальных распределений, и выведите формулы для Е- и М-шагов ЕМ-алгоритма.

Задача 3. Рассмотрим смесь двух одномерных нормальных распределений:

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

Дисперсии σ_1^2 и σ_2^2 известны.

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