Теория 3. Вариационный вывод

Курс: Байесовские методы в машинном обучении, 2020

Пусть $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$, $\boldsymbol{x}_n \in \mathbb{R}^D$ – независимая выборка из смеси распределений Стьюдента

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} w_k \mathcal{T}(\boldsymbol{x}|\boldsymbol{\mu}_k, \Sigma_k, \nu), \quad w_k \geq 0, \; \sum_j w_j = 1. \tag{1}$$

Рассмотрим следующую вероятностную модель со скрытыми переменными:

$$p(X, T, Z|\boldsymbol{w}, \boldsymbol{\mu}, \Sigma, \nu) = \prod_{n,k=1}^{N,K} \Big[w_k \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k, \Sigma_k/z_n) \mathcal{G}(z_n|\nu/2, \nu/2) \Big]^{t_{nk}}. \tag{2}$$

Здесь $t_{nk} \in \{0, 1\}$, $\sum_j t_{nj} = 1$ обозначает принадлежность n-го объекта k-ой компоненте смеси. Очевидно, что неполное правдоподобие $p(X|\boldsymbol{w}, \boldsymbol{\mu}, \Sigma, \nu)$ для модели (2) совпадает с правдоподобием выборки X для смеси (1). Поэтому оценки максимального правдоподобия $w_{ML,k}, \mu_{ML,k}, \Sigma_{ML,k}$ для смеси (1) можно искать с помощью вариационного EM-алгоритма для модели (2), в котором на E-шаге апостериорное распределение приближается в семействе

$$q_T(T) q_Z(Z) \approx p(T, Z|X, \boldsymbol{w}, \boldsymbol{\mu}, \Sigma, \nu).$$

Для выполнения задания требуется:

1. Выписать формулы пересчёта для компонент вариационного приближения $q_T(T)$ и $q_Z(Z)$;
2. Выписать формулы пересчёта параметров w_k, μ_k, Σ_k на M-шаге. Убедиться, что эти формулы переходят в соответствующие формулы с семинара по EM-алгоритму для случая $K = 1$;
3. Расписать функционал $\mathcal{L}(q, \boldsymbol{w}, \boldsymbol{\mu}, \Sigma)$ – нижнюю оценку на $\log p(X|\boldsymbol{w}, \boldsymbol{\mu}, \Sigma, \nu)$;
4. Найти формулы для статистик распределений $q_T(T)$ и $q_Z(Z)$, требуемых в предыдущих трёх пунктах.