ВМК МГУ

Семинар 6. EM-алгоритм

Курс: Байесовские методы в машинном обучении, 2016

1. Пусть $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ – независимая выборка из смеси нормальных распределений

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} w_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \Sigma_k),\ \sum_k w_k = 1,\ w_k \geq 0.$$

Вывести формулы EM-алгоритма для поиска оценок максимального правдоподобия $\boldsymbol{w}_{ML}, \boldsymbol{\mu}_{k,ML}, \Sigma_{k,ML}$.

2. Пусть $x_1, \ldots, x_N$ – независимая выборка из смеси $p(x) = \gamma p_1(x) + (1-\gamma)p_2(x)$, где $x_n \in \{1, 2, 3\}$ и

$$p_1: \begin{array}{ccc} 1 & 2 & 3 \\ \alpha & 1-\alpha & 0 \end{array} \qquad p_2: \begin{array}{ccc} 1 & 2 & 3 \\ 0 & 1-\beta & \beta \end{array}$$

В наблюдаемой выборке X содержится 30 единиц, 20 двоек и 60 троек. Требуется провести первые две итерации EM-алгоритма для разделения данной смеси из начального приближения $\gamma_0 = \alpha_0 = \beta_0 = \frac{1}{2}$.

3. Доказать, что многомерное распределение Стьюдента с плотностью

$$\mathcal{T}(\boldsymbol{x}|\boldsymbol{\mu}, \Sigma, \nu) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}^d\sqrt{\det\Sigma}\left[1 + \frac{1}{\nu}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]^{\frac{\nu+d}{2}}}, \quad \boldsymbol{x} \in \mathbb{R}^d$$

может быть представлено как

$$\int_0^{+\infty} \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, z^{-1}\Sigma)\mathcal{G}(z \mid \nu/2, \nu/2)dz.$$

4. Пусть $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ – независимая выборка из $\mathcal{T}(\boldsymbol{x}|\boldsymbol{\mu}, \Sigma, \nu)$. Вывести формулы EM-алгоритма для поиска оценок максимального правдоподобия $\boldsymbol{\mu}_{ML}, \Sigma_{ML}, \nu_{ML}$.

5. Пусть $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ – независимая выборка из смеси распределений Стьюдента

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} w_k \mathcal{T}(\boldsymbol{x}|\boldsymbol{\mu}_k, \Sigma_k, \nu_k),\ \sum_k w_k = 1,\ w_k \geq 0.$$

Вывести формулы EM-алгоритма для поиска оценок максимального правдоподобия $\boldsymbol{w}_{ML}, \boldsymbol{\mu}_{k,ML}, \Sigma_{k,ML}, \nu_{k,ML}$.

6. Рассматривается модель регрессии релевантных векторов. Вывести формулы EM-алгоритма для решения задачи максимизации обоснованности:

$$p(\boldsymbol{t}|X, A, \beta) = \int \mathcal{N}(\boldsymbol{t}|X\boldsymbol{w}, \beta^{-1}I)\mathcal{N}(\boldsymbol{w}|\boldsymbol{0}, A)d\boldsymbol{w} \to \max_{A,\beta},$$

где $A = \mathrm{diag}(\alpha_1, \ldots, \alpha_d)$.