

Домашнее задание по материалу 4-го семинара.

ММП, осень 2012–2013

12 октября

Важно: на проверочной работе обязательно будет задача в духе «выпишите E и M шаги общего EM-алгоритма для заданной вероятностной модели» для какой-нибудь простой модели, отличной от смеси. Поэтому очень внимательно разберитесь с EM-алгоритмом в общем виде (задачи 2–3). Также полезно научиться работать с дивергенцией.

1) Дано два многомерных нормальных распределения с единичными ковариационными матрицами: $P(x)$ с центром в начале координат и $Q(x)$ со средним μ_Q . Докажите, что

$$KL(Q\|P) = \|\mu_Q\|^2/2,$$

где $KL(\cdot\|\cdot)$ — KL-дивергенция, а $\|\cdot\|$ — обычная евклидова норма.

2) Пусть дана вероятностная модель с наблюдаемыми переменными $X = \{\mathbf{x}_1, \dots, \mathbf{x}_\ell\}$, $\mathbf{x}_i \in \mathbb{R}^n$, соответствующими им скрытыми переменными $Z = \{z_1, \dots, z_\ell\}$ и вектором параметров $\theta = \{\mu_k, \Sigma_k, \pi_k, k = 1, \dots, K\}$, рассмотренная на 4-ом семинаре. Скрытые переменные на этот раз будут принимать скалярные значения из множества $\{1, \dots, K\}$ (а не бинарные вектора с одной единицей, как на семинаре). По-прежнему,

$$p(z = k) = \pi_k \geq 0, \quad k = 1, \dots, K, \quad \sum_k \pi_k = 1;$$

$$p(\mathbf{x}|z = k) = \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k), \quad \mathbf{x} \in \mathbb{R}^n.$$

Убедитесь, что EM-алгоритм для смеси распределений, изложенный на лекции, получается применением общего EM-алгоритма для максимизации правдоподобия наблюдаемых переменных в этой модели. (Аккуратно выпишите E и M шаги общего EM-алгоритма).

3) Предположим, что в общем EM-алгоритме из прошлой задачи на E-шаге мы минимизируем дивергенцию $KL(q\|p)$ по семейству вырожденных распределений q (распределение вырождено, если оно всю вероятностную меру сосредотачивает на одной точке). Также предположим, что ковариационные матрицы всех компонент совпадают и равны единичной. Как будут выглядеть в этом случае EM итерации?

Указания: в этом случае мы получим так называемый алгоритм «K-средних», который будет рассказан на одной из следующих лекций. В отличие от рассмотренного ранее EM-алгоритма расщепления смеси, в этом случае а) каждая точка обучающей выборки будет жестко приписываться одной из компонент, б) средние каждой

из компонент на М-шаге будут пересчитываться по точкам, отнесенным к ним, в) веса смеси вычисляются как доля точек, отнесенных к компонентам.