

EM-алгоритм

Ветров,
Кропотов,
Осокин

EM для
гауссовской
смеси

EM в общем виде

EM как
покоординатный
спуск

EM-алгоритм

Д. П. Ветров¹ Д. А. Кропотов² А. А. Осокин¹

¹МГУ, ВМиК, каф. ММП

²ВЦ РАН

Спецкурс «Байесовские методы машинного обучения»

План лекции

EM-алгоритм

Ветров,
Кропотов,
Осокин

EM для
гауссовской
смеси

EM в общем виде

EM как
покоординатный
спуск

- 1 EM для гауссовской смеси
 - Нормальное распределение
 - Постановка задачи
 - Вывод формул EM-алгоритма
 - EM-алгоритм
- 2 EM в общем виде
 - Смесь нормальных распределений
- 3 EM как покоординатный спуск

Нормальное распределение

EM-алгоритм

Ветров,
Кропотов,
Осокин

EM для
гауссовской
смеси

Нормальное
распределение

Постановка
задачи

Вывод формул
EM-алгоритма
EM-алгоритм

EM в общем виде

EM как
покоординатный
спуск

- Многомерное нормальное распределение имеет вид

$$X \sim \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}, \Sigma) = \frac{1}{\sqrt{2\pi}^n \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)$$

где $\boldsymbol{\mu} = \mathbb{E}X$, $\Sigma = \mathbb{E}(X - \boldsymbol{\mu})(X - \boldsymbol{\mu})^T$ — вектор математических ожиданий каждой из n компонент и матрица ковариаций соответственно

- Смесь нормальных распределений

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

Смесь нормальных распределений

EM-алгоритм

Ветров,
Кропотов,
Осокин

EM для
гауссовской
смеси

Нормальное
распределение

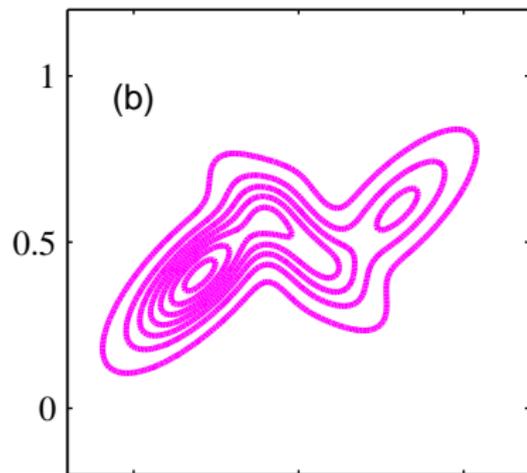
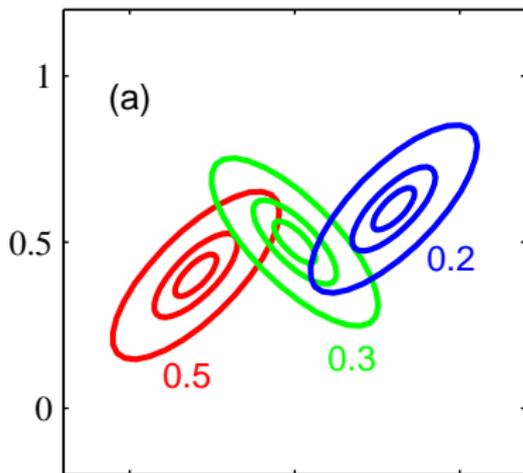
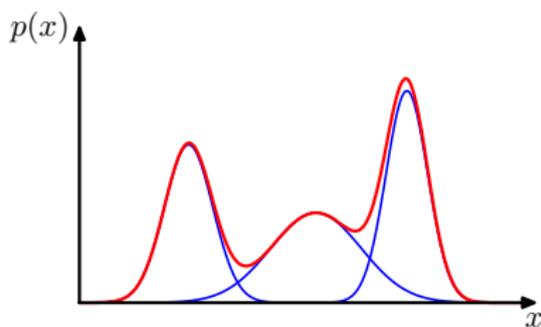
Постановка
задачи

Вывод формул
EM-алгоритма

EM-алгоритм

EM в общем виде

EM как
покоординатный
спуск



Смесь нормальных распределений

EM-алгоритм

Ветров,
Кропотов,
Осокин

EM для
гауссовской
смеси

Нормальное
распределение

Постановка
задачи

Вывод формул
EM-алгоритма
EM-алгоритм

EM в общем виде

EM как
покоординатный
спуск

- Правило суммы и произведения:

$$p(\mathbf{x}) = \sum_{k=1}^K p(k)p(\mathbf{x} | k) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k).$$

- Апостериорные вероятности:

$$\gamma_k(\mathbf{x}) = p(k | \mathbf{x}) = \frac{p(k)p(\mathbf{x} | k)}{\sum_l p(l)p(\mathbf{x} | l)} = \frac{\pi_k \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_l \pi_l \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)}.$$

Постановка задачи

EM-алгоритм

Ветров,
Кропотов,
Осокин

EM для
гауссовской
смеси

Нормальное
распределение

Постановка
задачи

Вывод формул
EM-алгоритма
EM-алгоритм

EM в общем виде

EM как
покоординатный
спуск

- Дана выборка X .
- Применить М.М.П.:

$$\ln p(X | \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\}.$$

- Получается сложная оптимизационная задача с сингулярностями. (E.g., если компонента превращается в одну точку).

Вывод формул.

EM-алгоритм

Ветров,
Кропотов,
Осокин

EM для
гауссовской
смеси

Нормальное
распределение
Постановка
задачи

Вывод формул
EM-алгоритма

EM-алгоритм

EM в общем виде

EM как
покоординатный
спуск

EM-алгоритм

EM-алгоритм

Ветров,
Кропотов,
Осокин

EM для
гауссовской
смеси

Нормальное
распределение
Постановка
задачи

Вывод формул
EM-алгоритма

EM-алгоритм

EM в общем виде

EM как
покоординатный
спуск

- 1 Инициализировать μ_k, Σ_k, π_k .
- 2 **E-шаг.** Вычислить апостериорные вероятности:

$$\gamma_{nk} = \frac{\pi_k \mathcal{N}(\mathbf{x}_n | \mu_k, \Sigma_k)}{\sum_l \pi_l \mathcal{N}(\mathbf{x}_n | \mu_l, \Sigma_l)}.$$

- 3 **M-шаг.** Пересчет параметров.

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma_{nk} \mathbf{x}_n,$$

$$\Sigma_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma_{nk} (\mathbf{x}_n - \mu_k^{new})(\mathbf{x}_n - \mu_k^{new})^T,$$

$$\pi_k = \frac{N_k}{N}, \quad N_k = \sum_{n=1}^N \gamma_{nk}.$$

- 4 Вычислить логарифм правдоподобия, и, если не останов, то шаг 2.

Пример

EM-алгоритм

Ветров,
Кропотов,
Осокин

EM для
гауссовской
смеси

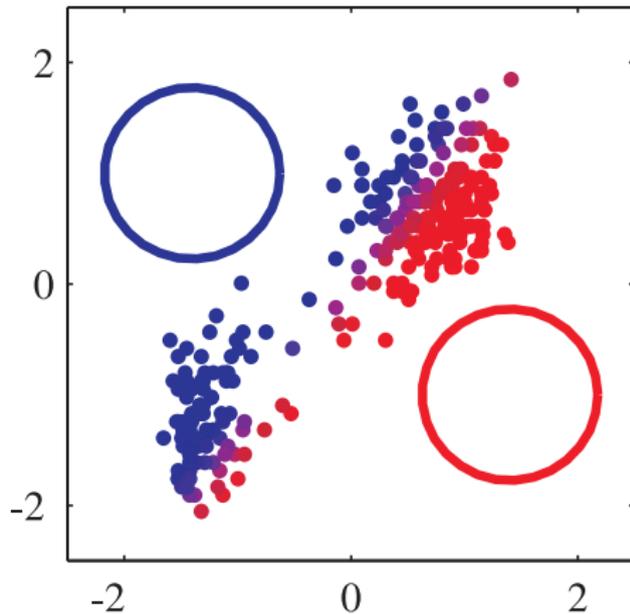
Нормальное
распределение
Постановка
задачи

Вывод формул
EM-алгоритма

EM-алгоритм

EM в общем виде

EM как
покоординатный
спуск



Пример

EM-алгоритм

Ветров,
Кропотов,
Осокин

EM для
гауссовской
смеси

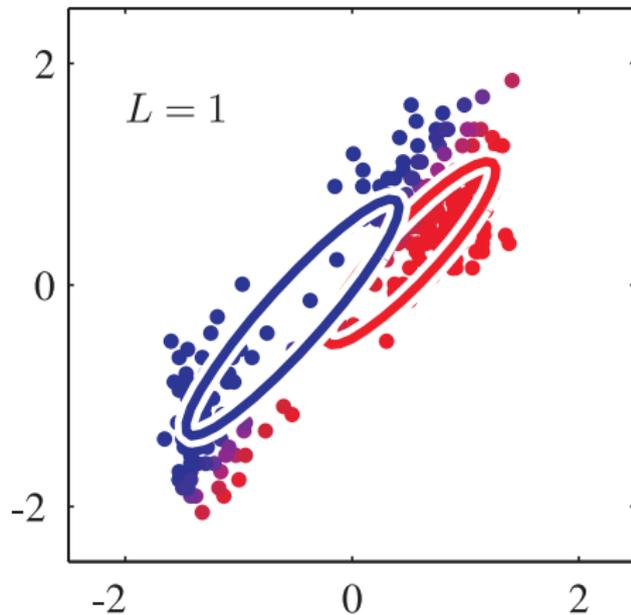
Нормальное
распределение
Постановка
задачи

Вывод формул
EM-алгоритма

EM-алгоритм

EM в общем виде

EM как
покоординатный
спуск



Пример

EM-алгоритм

Ветров,
Кропотов,
Осокин

EM для
гауссовской
смеси

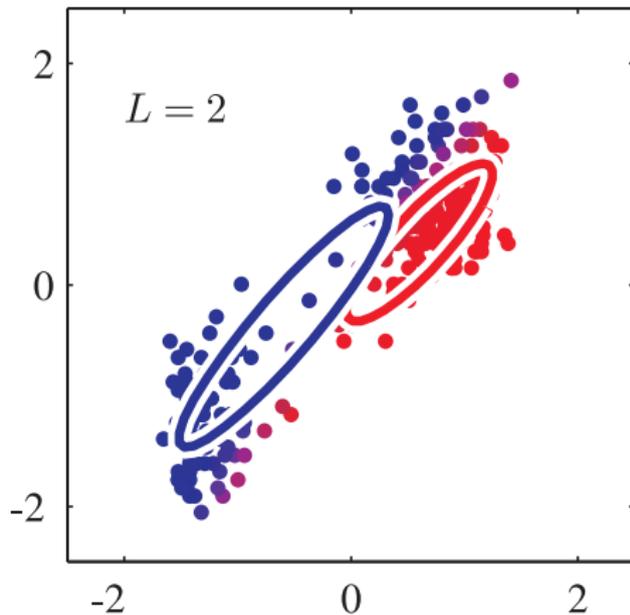
Нормальное
распределение
Постановка
задачи

Вывод формул
EM-алгоритма

EM-алгоритм

EM в общем виде

EM как
покоординатный
спуск



Пример

EM-алгоритм

Ветров,
Кропотов,
Осокин

EM для
гауссовской
смеси

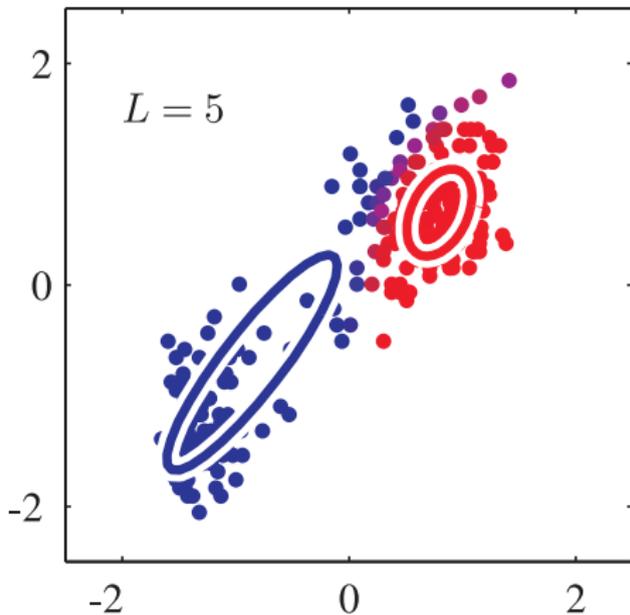
Нормальное
распределение
Постановка
задачи

Вывод формул
EM-алгоритма

EM-алгоритм

EM в общем виде

EM как
покоординатный
спуск



Пример

EM-алгоритм

Ветров,
Кропотов,
Осокин

EM для
гауссовской
смеси

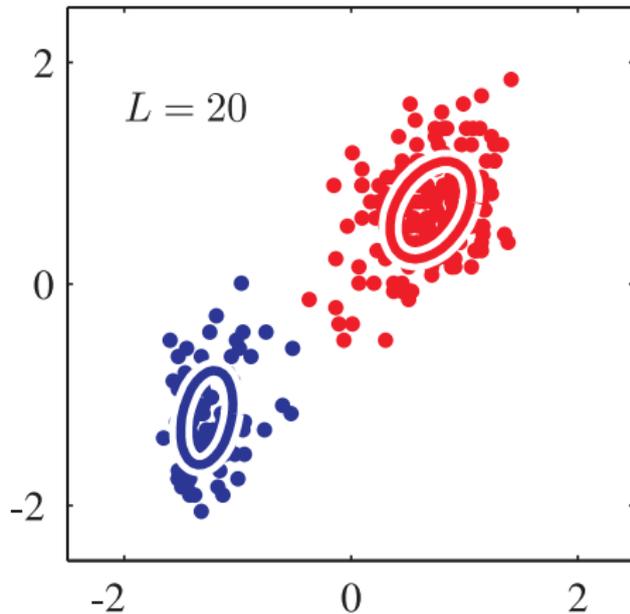
Нормальное
распределение
Постановка
задачи

Вывод формул
EM-алгоритма

EM-алгоритм

EM в общем виде

EM как
покоординатный
спуск



EM-алгоритм

EM-алгоритм

Ветров,
Кропотов,
Осокин

EM для
гауссовской
смеси

EM в общем виде

Смесь
нормальных
распределений

EM как
покоординатный
спуск

- X – наблюдаемые переменные; Z – не наблюдаемые переменные; θ – параметры
- $p(X | \theta)$ – неполное правдоподобие; $p(X, Z | \theta)$ – полное правдоподобие
- Соотношение полного и неполного правдоподобия:

$$\ln p(X | \theta) = \ln \left\{ \sum_Z p(X, Z | \theta) \right\}.$$

- Сумма внутри логарифма!
- О скрытых переменных известно распределение $p(Z | X, \theta^{old})$
- Вместо полного правдоподобия возьмем его М.О. по апостериорному распределению.

EM-алгоритм

EM-алгоритм

Ветров,
Кропотов,
Осокин

EM для
гауссовской
смеси

EM в общем виде

Смесь
нормальных
распределений

EM как
покоординатный
спуск

- 1 Инициализировать θ^{old} .
- 2 **E-шаг.** Вычислить апостериорные вероятности:

$$p(Z | X, \theta^{old})$$

- 3 **M-шаг.** Пересчет параметров.

$$\theta^{new} = \arg \max_{\theta} Q(\theta, \theta^{old}),$$

где

$$Q(\theta, \theta^{old}) = \sum_Z p(Z | X, \theta^{old}) \ln p(X, Z | \theta)$$

- 4 Вычислить логарифм правдоподобия, и, если не останов, то шаг 2.

$$\theta^{old} := \theta^{new}.$$

Смесь нормальных распределений.

EM-алгоритм

Ветров,
Кропотов,
Осокин

EM для
гауссовской
смеси

EM в общем виде

Смесь
нормальных
распределений

EM как
покоординатный
спуск

EM как покоординатный спуск

EM-алгоритм

Ветров,
Кропотов,
Осокин

EM для
гауссовской
смеси

EM в общем виде

EM как
покоординатный
спуск

$$p(X | \theta) = \sum_Z p(X, Z | \theta)$$

Заметим, что

$$p(X | \theta) = \mathcal{L}(q, \theta) + KL(q(Z) \| p(Z | X, \theta)),$$

где

$$\mathcal{L}(q, \theta) = \sum_Z q(Z) \ln \left\{ \frac{p(X, Z | \theta)}{q(Z)} \right\},$$

$$KL(q(Z) \| p(Z | X, \theta)) = - \sum_Z q(Z) \ln \left\{ \frac{p(Z | X, \theta)}{q(Z)} \right\}.$$

Это верно, поскольку

$$\ln p(X, Z | \theta) = \ln p(Z | X, \theta) + \ln p(X | \theta).$$

KL-дивергенция

EM-алгоритм

Ветров,
Кропотов,
Осокин

EM для
гауссовской
смеси

EM в общем виде

EM как
покоординатный
спуск

$$KL(q(Z)\|p(Z)) = - \sum_Z q(Z) \ln \left\{ \frac{p(Z)}{q(Z)} \right\}.$$

- $KL(q(Z)\|p(Z)) \geq 0$
- $KL(q(Z)\|p(Z)) = 0 \iff q \equiv p.$
- KL несимметрична

Нижняя граница

EM-алгоритм

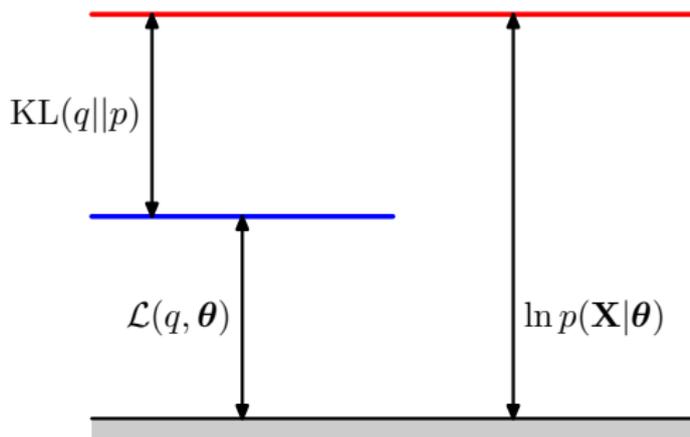
Ветров,
Кропотов,
Осокин

EM для
гауссовской
смеси

EM в общем виде

EM как
покоординатный
спуск

$$p(X | \theta) = \mathcal{L}(q, \theta) + KL(q(Z) || p(Z | X, \theta)),$$



$KL \geq 0$, а значит \mathcal{L} – нижняя граница неполного правдоподобия $p(X | \theta)$.

E-шаг

EM-алгоритм

Ветров,
Кропотов,
Осокин

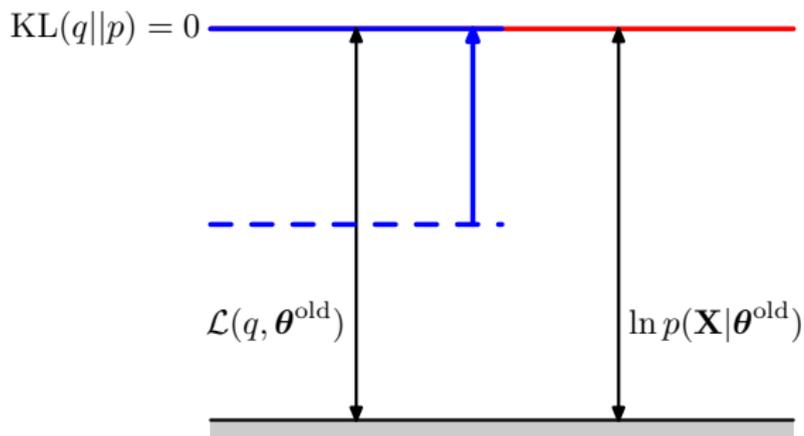
EM для
гауссовской
смеси

EM в общем виде

EM как
покоординатный
спуск

$$p(\mathbf{X} | \theta) = \mathcal{L}(q, \theta) + KL(q(Z) || p(Z | \mathbf{X}, \theta)),$$

Максимизируем $\mathcal{L}(q, \theta)$ по q за счет минимизации KL .



M-шаг

EM-алгоритм

Ветров,
Кропотов,
Осокин

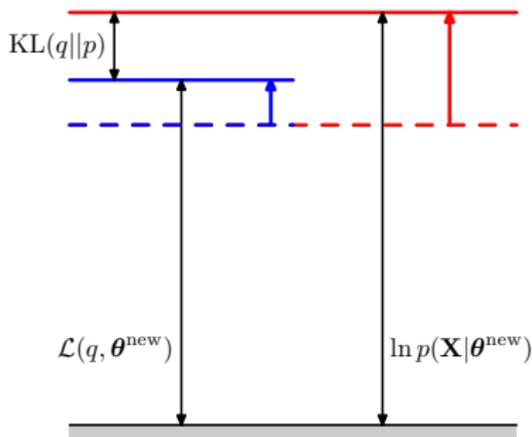
EM для
гауссовской
смеси

EM в общем виде

EM как
покоординатный
спуск

$$p(X | \theta) = \mathcal{L}(q, \theta) + KL(q(Z) || p(Z | X, \theta)),$$

Максимизируем $\mathcal{L}(q, \theta)$ по θ .



Если $q(Z) = p(Z | X, \theta^{old})$, то

$$\mathcal{L}(q, \theta) = \sum_Z p(Z | X, \theta^{old}) \ln p(X, Z | \theta) - \sum_Z p(Z | X, \theta^{old}) \ln p(Z | X,$$

$$= \mathcal{Q}(\theta, \theta^{old}) + \text{const.}$$

Варианты E-шага

EM-алгоритм

Ветров,
Кропотов,
Осокин

EM для
гауссовской
смеси

EM в общем виде

EM как
покоординатный
спуск

- $p(Z | X, \theta^{old})$ точно вычислимо
- $p(Z | X, \theta^{old}) = \prod_i q(Z_i)$ – вариационный подход (будет на следующих лекциях)
- $p(Z | X, \theta^{old}) = \delta(\arg \max_Z p(Z | X, \theta^{old}))$ – жесткое приписывание классам. Например, K средних.