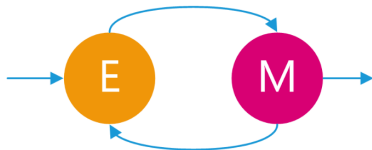


# ЕМ алгоритм

Виктор Китов

[victorkitov.github.io](https://victorkitov.github.io)

Курс поддержан  
фондом  
'Интеллект'



Победитель  
конкурса VK среди  
курсов по IT



# Содержание

- 1 Неравенство Йенсена
- 2 ЕМ-алгоритм
- 3 ЕМ с регуляризацией
- 4 Независимые наблюдения  $(x_n, z_n)$

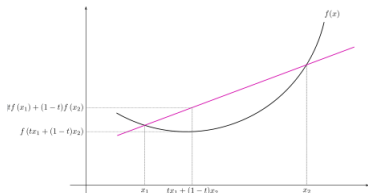
## Строго выпуклые функции

- Множество  $X$  выпукло, если  $\forall x, y \in X, \forall \alpha \in (0, 1)$  :

$$\alpha x + (1 - \alpha)y \in X$$

- Функция  $f(x)$  строго выпукла на выпуклом  $X$ , если  $\forall \alpha \in (0, 1), \forall x_1 \neq x_2 \in X$ :

$$f(\alpha x_1 + (1 - \alpha)x_2) < \alpha f(x_1) + (1 - \alpha)f(x_2)$$



- Что можно сказать о минимумах выпуклых/строго выпуклых ф-ций и достаточном условии минимума?

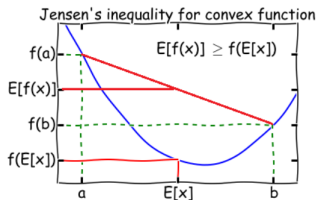
## Признаки и свойства<sup>1</sup>

- $f(x)$  строго выпукла  $\Leftrightarrow$  она всегда выше касательной

$$f(y) > f(x) + \nabla f(x)^T (y - x) \quad \forall y \neq x \in X$$

- Если  $\nabla^2 f(x) \succ 0 \quad \forall x \in X$ , то  $f(x)$  - строго выпукла на  $X$ .
- Если  $f(x)$  - строго выпукла, то выполнено нер-во Йенсена

$$\mathbb{E}[f(X)] > f(\mathbb{E}X) \quad \forall X \stackrel{\text{п.в.}}{\neq} \text{const}$$



<sup>1</sup>Докажите утверждения. Верны ли они в обратную сторону?

## Доказательство неравенства Йенсена

Для строго выпуклой  $f(x)$ :

$$f(x) > f(y) + \nabla f(y)^T (x - y)$$

в частности, для не константной  $X$  подставим  $x = X$  и  $y = \mathbb{E}X$

$$f(X) > f(\mathbb{E}X) + \nabla f(\mathbb{E}X)^T (X - \mathbb{E}X)$$

$$\mathbb{E} : \quad \mathbb{E}f(X) > f(\mathbb{E}X) + \nabla f(\mathbb{E}X)^T (\mathbb{E}X - \mathbb{E}X) = f(\mathbb{E}X)$$

Для  $X \stackrel{\text{п.в.}}{=} \mathbb{E}X$ :

$$f(X) = f(\mathbb{E}X)$$

$$\mathbb{E} : \quad \mathbb{E}f(X) = \mathbb{E}f(\mathbb{E}X) = f(\mathbb{E}X)$$

# Содержание

- 1 Неравенство Йенсена
- 2 **EM-алгоритм**
- 3 EM с регуляризацией
- 4 Независимые наблюдения  $(x_n, z_n)$

## Вероятностная модель

Рассмотрим вероятностную модель с наблюдаемыми переменными  $x$  и ненаблюдаемыми (латентными) переменными  $z$ .

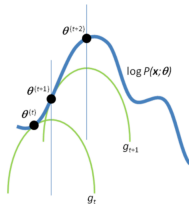
- обозначим  $X = [x_1, x_2, \dots, x_N]$ , и  $Z = [z_1, z_2, \dots, z_M]$ .

Для нахождения  $\hat{\theta}$  решим:

$$L(\theta) = \ln p_{\theta}(X) = \ln \sum_Z p_{\theta}(X, Z) \rightarrow \max_{\theta}$$

- Решение  $p_{\theta}(X, Z) \rightarrow \max_{\theta}$  не применимо, т.к. не знаем  $Z$ .
- Оценим распределение  $Z \sim q(Z)$ , зная  $X$ , и решим  $\mathbb{E}_Z p_{\theta}(X, Z) \rightarrow \max_{\theta}$ .
- Повторять до сходимости (ЕМ алгоритм):
  - Е шаг: оценить, как распределено  $Z$  при  $\hat{\theta}$
  - М шаг: максимизировать  $\ln p_{\theta}(X, Z)$ , усредненное по вариантам  $Z$ .

# Общая идея ЕМ алгоритма



$L(\theta) \geq G(q(Z), \theta) \quad \forall q(Z), \forall \theta$  ( $G$ -нижняя граница  $L \quad \forall \theta$  и  $\forall q(Z)$ )

- Инициализировать  $\hat{\theta}_0$  случайно,  $t = 0$
- Повторять до сходимости:
  1. Выбрать  $q(Z|\hat{\theta}_t)$  так, что  $L(\hat{\theta}_t) = G(q(Z), \hat{\theta}_t)$
  2.  $\hat{\theta}_{t+1} = \arg \max_{\theta} G(q(Z|\hat{\theta}_t), \theta)$
  3.  $t = t + 1$



# Комментарии

- Е-шаг:  $G(q(Z), \hat{\theta}_t) = \arg \max_{p(Z)} G(p(Z), \hat{\theta}_t)$
- М-шаг:  $\hat{\theta}_{t+1} = \arg \max_{\theta} G(q(Z), \theta)$
- ЕМ алгоритм - метод покоординатного подъема нижней границы  $G(p(Z), \theta)$ .
- $L(\hat{\theta}_t)$  сходится, т.к.
  - 1  $L(\hat{\theta}_t) = G(q(Z), \hat{\theta}_t) \leq G(q(Z), \hat{\theta}_{t+1}) \leq L(\hat{\theta}_{t+1}) \Rightarrow \left\{ L(\hat{\theta}_t) \right\} \uparrow$
  - 2  $\left\{ L(\hat{\theta}_t) \right\}$  ограничена сверху, т.к.  $L(\theta) = \ln p(X|\theta) \leq \ln 1$

## Вывод нижней оценки

Пусть  $q(Z)$  - некоторое распределение над  $Z$ ,  $q(Z) \geq 0$ ,  $\sum_Z q(Z) = 1$ . Тогда

$$\begin{aligned} L(\theta) &= \ln p_\theta(X) = \ln \sum_Z p_\theta(X, Z) \\ &= \ln \sum_Z q(Z) \frac{p_\theta(X, Z)}{q(Z)} \end{aligned} \quad (1)$$

$$\geq \sum_Z q(Z) \ln \frac{p_\theta(X, Z)}{q(Z)} = G(q(Z), \theta) \quad (2)$$

Использовали неравенство Йенсена  $f(\mathbb{E}U) \geq \mathbb{E}(fU) \forall$  сл.вел.  $U$  и вогнутой  $f$ .

- 1  $f(x) = \ln x$  вогнута, т.к.  $(\ln x)'' = -\frac{1}{x^2} < 0$
- 2 сл. вел.  $U: p\left(U = \frac{p(X, Z, \theta)}{q(Z)}\right) = q(Z)$  для всевозможных  $Z$ .

Е-шаг: делаем нижнюю грань точной при  $\hat{\theta}_t$

- Неравенство Йенсена:  $f(\mathbb{E}U) \geq \mathbb{E}(fU)$ , при этом  $f(\mathbb{E}U) = \mathbb{E}(fU) \iff U \stackrel{\text{п.в.}}{=} c = \text{const}$ :
- $L(\hat{\theta}_t) = G(q(Z), \hat{\theta}_t)$  при

$$U = \frac{p_{\hat{\theta}_t}(X, Z)}{q(Z)} = c \quad \forall Z$$

$$cq(Z) = p_{\hat{\theta}_t}(X, Z)$$

$$c \sum_Z q(Z) = \sum_Z p_{\hat{\theta}_t}(X, Z)$$

$$c = p_{\hat{\theta}_t}(X)$$

$$q(Z) = \frac{p_{\hat{\theta}_t}(X, Z)}{p_{\hat{\theta}_t}(X)} = p_{\hat{\theta}_t}(Z|X)$$

М-шаг: усредненный  $\log(\text{правдоподобия}) \rightarrow \max$

М-шаг: усредненный  $\log(\text{правдоподобия}) \rightarrow \max$

$$\begin{aligned}
 \hat{\theta}_{t+1} &= \arg \max_{\theta} \left\{ \sum_Z q(Z) \ln \frac{p_{\theta}(X, Z)}{q(Z)} \right\} \\
 &= \arg \max_{\theta} \left\{ \sum_Z q(Z) \ln p_{\theta}(X, Z) - \overbrace{\sum_Z q(Z) \ln q(Z)}^{\text{const}(\theta)} \right\} \\
 &= \arg \max_{\theta} \left\{ \sum_Z q(Z) \ln p_{\theta}(X, Z) \right\} \\
 &= \arg \max_{\theta} \left\{ \mathbb{E}_{Z \sim q(Z)} \ln p_{\theta}(X, Z) \right\}
 \end{aligned}$$

Замечание: от  $\theta$  зависит лишь  $p_{\theta}(X, Z)$ ,  $q(Z) = p_{\hat{\theta}_t}(Z|X)$  - не зависит

# ЕМ алгоритм

ВХОД:

выборка  $X = [x_1, \dots, x_N]$ , критерий сходимости

АЛГОРИТМ:

Инициализировать  $t = 0$ ,  $\theta_0$  - случайно

ПОВТОРЯТЬ до сходимости:

Е-шаг: уточнить распределение

над латентными переменными:

$$q(Z) = p(Z|X, \hat{\theta}_t)$$

М-шаг: уточнить параметры  $\theta$ :

$$\hat{\theta}_{t+1} = \arg \max_{\theta} \{ \sum_Z q(Z) \ln p(X, Z|\theta) \}$$

$$t = t + 1$$

ВЫХОД:  $\hat{\theta}_{t+1}$

## Комментарии по ЕМ алгоритму

- Возможные критерии сходимости:
  - $\|\hat{\theta}_{t+1} - \hat{\theta}_t\| < \varepsilon$
  - $L(\hat{\theta}_{t+1}) - L(\hat{\theta}_t) < \varepsilon$
  - $\# \text{итераций} > \text{порога}$
- ЕМ сходится к локальному оптимуму
  - можно перезапустить несколько раз из разных  $\hat{\theta}_0$  и выбрать лучшее решение
- Обобщенный ЕМ алгоритм (generalized EM, GEM)
  - для сходимости достаточно выбрать  $\hat{\theta}_{t+1}$  так, что

$$G(q(Z), \hat{\theta}_{t+1}) > G(q(Z), \hat{\theta}_t)$$

- например, сделать один шаг в оптимизации
- а не решать  $\hat{\theta}_{t+1} = \arg \max_{\theta} G(q(Z), \theta)$  точно.

# Содержание

- 1 Неравенство Йенсена
- 2 ЕМ-алгоритм
- 3 ЕМ с регуляризацией**
- 4 Независимые наблюдения  $(x_n, z_n)$

# ЕМ алгоритм с регуляризацией

- Добавим регуляризацию  $R(\theta)$  в задачу

$$L(\theta) = \ln p(X|\theta) - \lambda R(\theta) \rightarrow \max_{\theta}$$

- $R(\theta)$  штрафует сложность
- нужно вычитать, т.к.  $\ln p(X|\theta)$  максимизируется.
- Байесовская MAP оценка:

$$\ln p(X, \theta) = \ln p(X|\theta)p(\theta) = \ln p(X|\theta) + \underbrace{\ln p(\theta)}_{\lambda R(\theta)} \rightarrow \max_{\theta}$$

- Нижняя грань:

$$L(\theta) - \lambda R(\theta) \geq G(q(Z), \theta) - \lambda R(\theta) \quad \forall q(Z), \forall \theta$$



## ЕМ алгоритм с регуляризацией

- **Е-шаг:** не меняется (равенство из неравенства Йенсена)

$$q(Z) = p_{\hat{\theta}_t}(Z|X)$$

- **М-шаг:**

$$\hat{\theta} = \arg \max_{\theta} \{ \mathbb{E}_{Z \sim q(Z)} \ln p_{\theta}(X, Z) - \lambda R(\theta) \}$$

# Содержание

- 1 Неравенство Йенсена
- 2 EM-алгоритм
- 3 EM с регуляризацией
- 4 Независимые наблюдения  $(x_n, z_n)$

Е-шаг для независимых  $(x_n, z_n)$ 

- Рассмотрим частный случай независимых наблюдений  $\{(x_n, z_n)\}_{n=1}^N$ ,  $x_n$  - наблюдаемые,  $z_n$  - латентные
  - пример: смесь Гауссиан,  $z_n$ -#компоненты,  $x_n$ -реализация.
- Е-шаг становится:

$$q(Z) = p(Z|X, \theta) = p(z_1|x_1, \theta) \dots p(z_N|x_N, \theta) = q_1(z_1) \dots q_N(z_N)$$

$$q_n(z_n) = p(z_n|x_n, \theta)$$

## M-шаг для независимых $(x_n, z_n)$

Для независимых объектов  $(x_n, z_n)$ :

$$\begin{aligned}
 \sum_Z q(Z) \ln p(X, Z|\theta) &= \sum_{z_1, \dots, z_N} q_1(z_1) \dots q_N(z_N) \ln \prod_{n=1}^N p(x_n, z_n|\theta) \\
 &= \sum_{z_1, \dots, z_N} q_1(z_1) \dots q_N(z_N) \ln p(x_n, z_n|\theta) = \\
 &= \sum_{n=1}^N q_n(z_n) \ln p(x_n, z_n|\theta) \prod_{k \neq n} \left( \underbrace{\sum_{z_k} q_k(z_k)}_{=1} \right) \\
 &= \sum_{n=1}^N q_n(z_n) \ln p(x_n, z_n|\theta) \rightarrow \max_{\theta}
 \end{aligned}$$