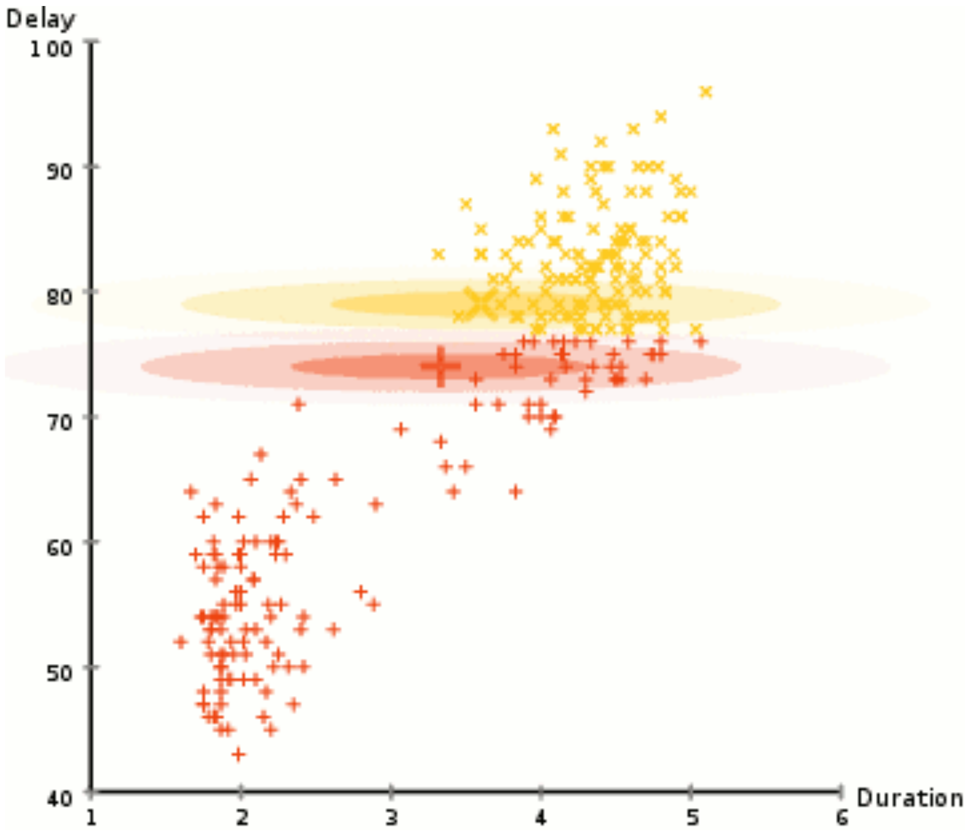
Expectation Maximization
(EM-алгоритм)

План

1. Как выглядит кластеризация с помощью ЕМ-алгоритма
2. Постановка задачи
3. Почему не решить «в лоб»
4. Описание ЕМ алгоритма
5. ЕМ-алгоритм в случае гауссовских распределений
6. Простое объяснение метода
7. Классическое объяснение метода
8. Для чего еще используют алгоритм

Как это выглядит



Здесь надо будет вставить анимацию:
https://en.wikipedia.org/wiki/File:EM_Clustering_of_Old_Faithful_data.gif

Постановка задачи

Модель порождения данных:

- Априорные вероятности кластеров - $w_1, \dots, w_K$
- Плотности распределения кластеров - $p_1(x), \dots, p_K(x)$
- Плотность распределения вектора признаков x:

$$p(x) = \sum_{j=1}^{K} w_j p_j(x)$$

Постановка задачи

Модель порождения данных:

- Априорные вероятности кластеров - $w_1, \dots, w_K$
- Плотности распределения кластеров - $p_1(x), \dots, p_K(x)$
- Плотность распределения вектора признаков x:

$$p(x) = \sum_{j=1}^{K} w_j p_j(x)$$

Что будем делать:

По выборке оценим параметры модели: $w_1, \dots, w_K$ и $p_1(x), \dots, p_K(x)$

Постановка задачи

Модель порождения данных:

- Априорные вероятности кластеров - $w_1, \dots, w_K$
- Плотности распределения кластеров - $p_1(x), \dots, p_K(x)$
- Плотность распределения вектора признаков x:

$$p(x) = \sum_{j=1}^{K} w_j p_j(x)$$

Что будем делать:

По выборке оценим параметры модели: $w_1, \dots, w_K$ и $p_1(x), \dots, p_K(x)$

Зачем:

Сможем оценивать вероятность принадлежности к кластеру

Постановка задачи: разделение смеси

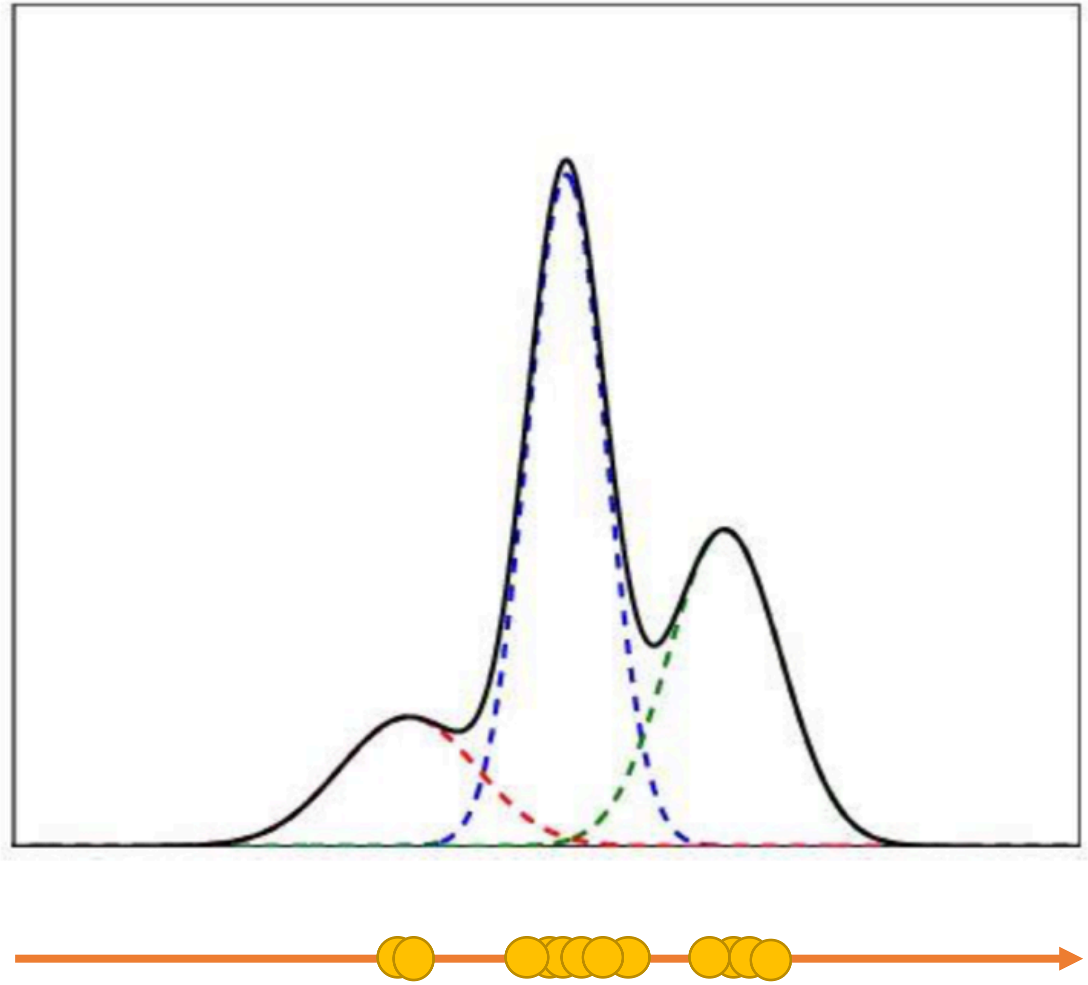
$$p(x) = \sum_{j=1}^{K} w_j p_j(x)$$

Оценить: $w_1, \dots, w_K$ и $p_1(x), \dots, p_K(x)$

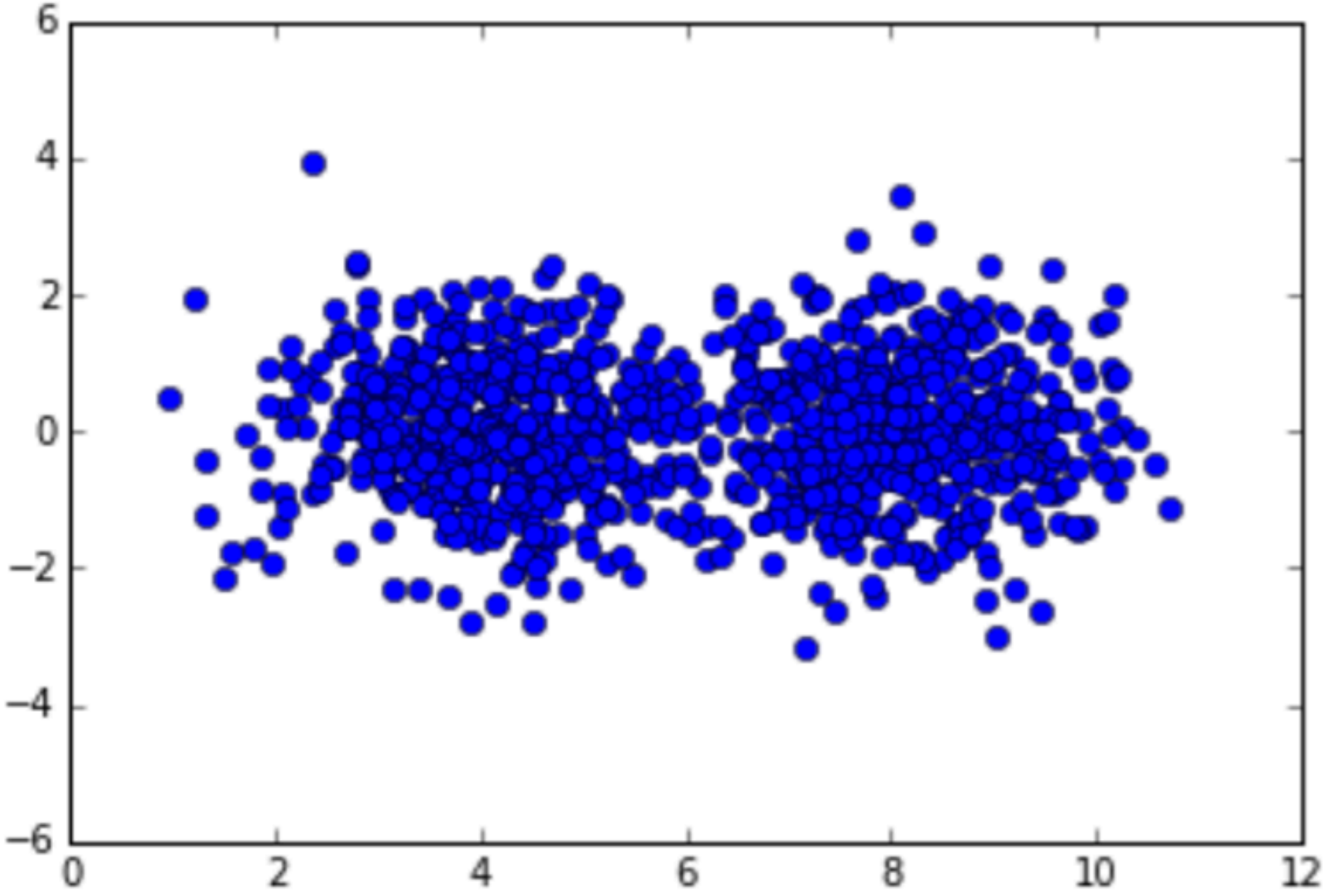
$$p_j(x) = \varphi(\theta_j; x)$$

Например, $p_j(x)$ - плотность нормального распределения (со своими параметрами для каждой компоненты)

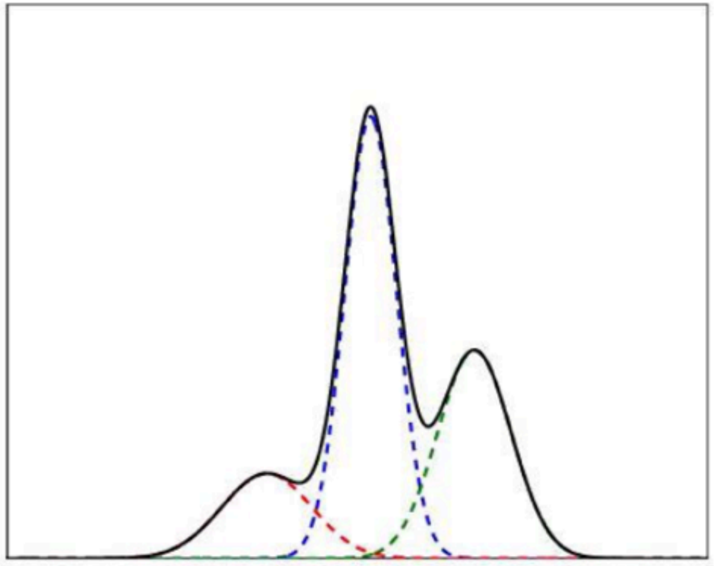
Как выглядит смесь распределений



Как выглядит смесь распределений



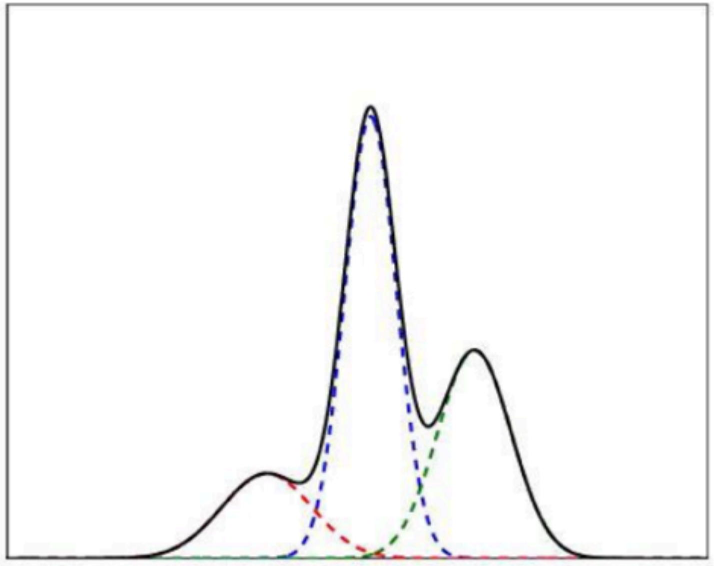
Почему не решить задачу «в лоб»



$$p(x) = \sum_{j=1}^{K} w_j p_j(x), \qquad p_j(x) = \varphi(\theta_j; x)$$

$$w, \theta = argmax_{\theta,w} \sum_{j=1}^{K} \ln p(x_i)$$

Почему не решить задачу «в лоб»



$$p(x) = \boxed{\sum_{j=1}^{K} w_j p_j(x)}, \qquad p_j(x) = \varphi(\theta_j; x)$$

$$w\,, \theta = argmax_{\theta,w} \sum_{j=1}^{K} \ln \boxed{p(x_i)}$$

ЕМ-алгоритм

$$p(x) = \sum_{j=1}^{K} w_j p_j(x), \qquad p_j(x) = \varphi(\theta_j; x)$$

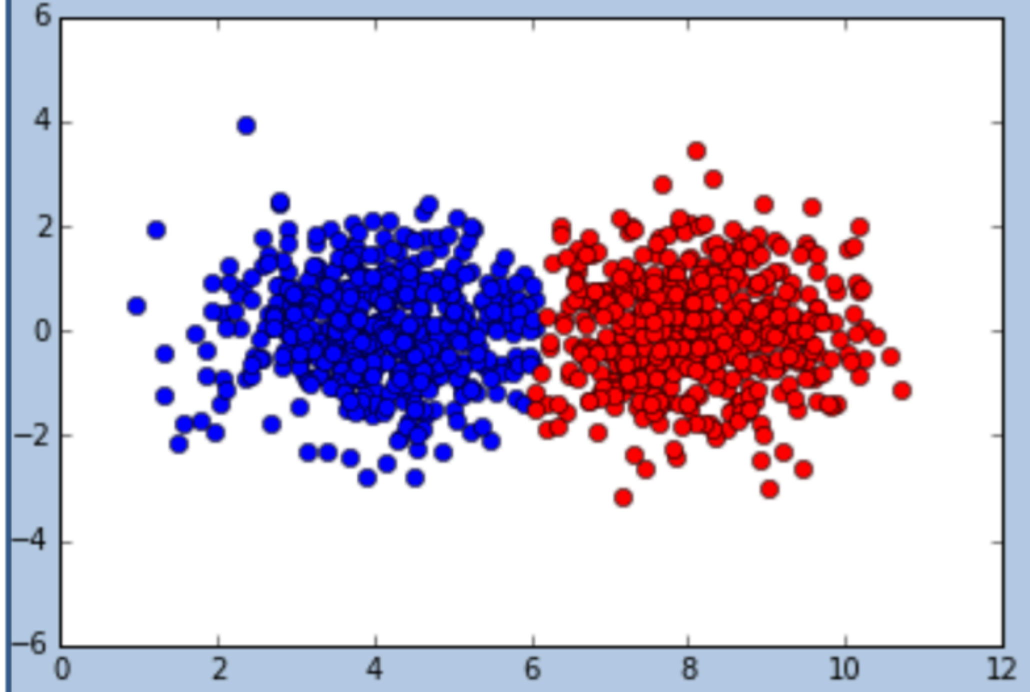
Е-шаг:

$$\mathrm{g_{ji}} = p(j|x_i) = \frac{w_j p_j(x_i)}{p(x_i)}$$

М-шаг:

$$w_j = \frac{1}{N}\sum_{i=1}^{N} g_{ji} \qquad \theta_j = argmax_\theta \sum_{i=1}^{N} g_{ji} \ln \varphi(\theta; x)$$

Пример: 2 кластера с гауссовской плотностью



Относим x_i к кластеру j, для которого больше $p(j|x_i) = g_{ij}$

$$p(x) = w_1 p_1(x) + w_2 p_2(x)$$

<u>Е-шаг:</u> $g_{ji} = p(j|x_i) = \frac{w_j p_j(x_i)}{p(x_i)}$

<u>М-шаг:</u>

$$w_j = \frac{1}{N} \sum_{i=1}^{N} g_{ji}$$

$$\mu_j = \frac{1}{N w_j} \sum_{i=1}^{N} g_{ij} x_i$$

$$\Sigma_j = \frac{1}{N w_j - 1} \sum_{i=1}^{N} g_{ij} (x_i - \mu_j)(x_i - \mu_j)^T$$

Простое объяснение EM-алгоритма

- Выбираем «скрытые переменные» таким образом, чтобы с ними было проще максимизировать правдоподобие
- Е-шаг:
  - Оцениваем скрытые переменные
- М-шаг:
  - Оцениваем $w_1, \dots, w_K$ и $p_1(x), \dots, p_K(x)$, считая скрытые переменные зафиксированными

Простое объяснение EM-алгоритма

- E-шаг:
  - Для задачи разделения смеси подходят $P(j|x_i)$
  - Расписав по формуле Байеса, получаем: $P(j|x_i) = \frac{w_j p_j(x_i)}{\sum_{k=1}^{K} w_k p_k(x_i)}$
- M-шаг:
  - Максимизируем правдоподобие по $w_1, \dots, w_K$ и $p_1(x), \dots, p_K(x)$, считая $P(j|x_i)$ константами
  - Если выписать производные по параметрам и приравнять к нулю, получаем:

$$w_j = \frac{1}{N}\sum_{i=1}^{N} g_{ji} \qquad \theta_j = argmax_\theta \sum_{i=1}^{N} g_{ji} \ln \varphi(\theta; x)$$

Классическое объяснение EM-алгоритма

$$L(\boldsymbol{\theta}; \mathbf{X}, \mathbf{Z}) = p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}),$$

$$L(\boldsymbol{\theta}; \mathbf{X}) = p(\mathbf{X}|\boldsymbol{\theta}) = \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$$

E-шаг: $$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \mathrm{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(t)}} \left[\log L(\boldsymbol{\theta}; \mathbf{X}, \mathbf{Z})\right]$$

M-шаг: $$\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$$

Какие еще задачи решаются с помощью ЕМ-алгоритма

- Оценка параметров в других вероятностных моделях (не только в смеси распределений)
- Восстановление плотности распределения
- Классификация

Резюме

1. Как выглядит кластеризация с помощью EM-алгоритма
2. Постановка задачи
3. Почему не решить «в лоб»
4. Описание EM алгоритма
5. EM-алгоритм в случае гауссовских распределений
6. Простое объяснение метода
7. Классическое объяснение метода
8. Для чего еще используют алгоритм