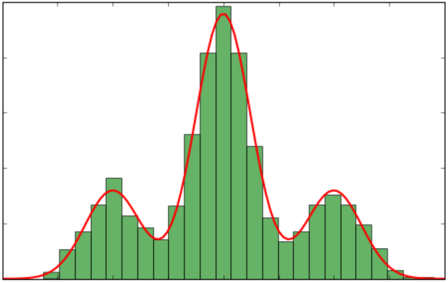
Смеси распределений

Виктор Китов

victorkitov.github.io



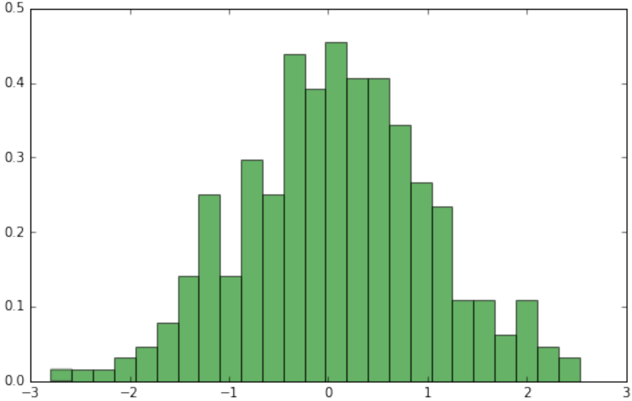
Курс поддержан фондом 'Интеллект'

Победитель конкурса VK среди курсов по IT

Содержание

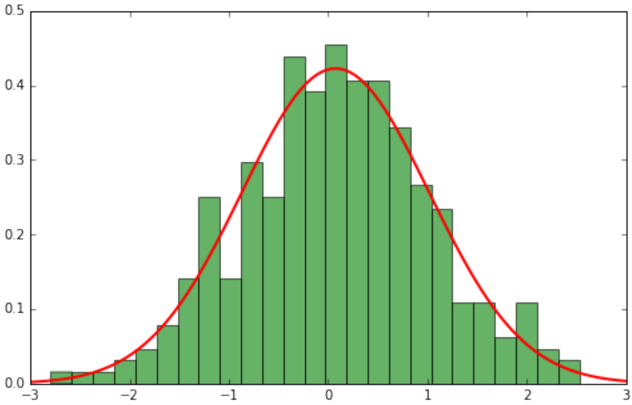
Выборочная плотность

Рассмотрим выборочную плотность:



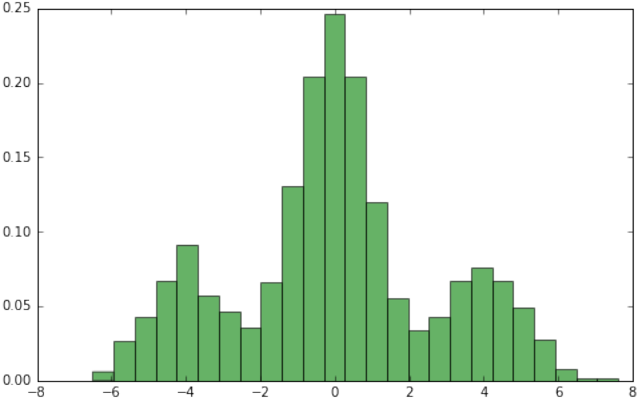
Параметрическая аппроксимация плотности

Возможна достаточно точная аппроксимация параметрическим семейством (Гауссовым)



Более общий вид параметрической плотности

Но как быть в таком случае?



Смеси распределений

$$p(x)=\sum_{k=1}^{K}\phi_k p(x;\theta_k)$$

- K - # компонент смеси
- $\phi_k,\ k=1,2,...K$ - вероятности компонент смеси, $\phi_k\ge 0,\ \sum_{k=1}^{K}\phi_k=1$
- $p(x;\theta_k)$ - индивидуальные плотности каждой компоненты
- Все параметры смеси распределений: $\Theta=\{\phi_k,\theta_k,\ k=1,2,...K\}$

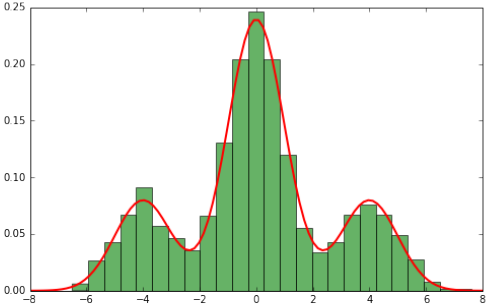
$p(x,\theta_k)$ м. быть из одинакового или разных параметрических семейств.

Смесь Гауссиан

Гауссово распределение - непрерывные сл. вел. на $(-\infty, +\infty)$.

$$p(x, \theta_k) = N(x, \mu_k, \Sigma_k),\ \theta_k = \{\mu_k, \Sigma_k\}$$

$$p(x) = \sum_{k=1}^{K} \phi_k N(x, \mu_k, \Sigma_k)$$



Смеси других распределений

Примеры распределений для моделирования:

- непрерывная на $(-\infty, +\infty)$
- непрерывная на $[a, \infty)$
- непрерывная на $[a, b]$
- дискретное на $[a, \infty)$
- дискретное на $[a, b]$

Смеси других распределений

Примеры распределений для моделирования:

- непрерывная на $(-\infty, +\infty)$
  - Гаусса, Лапласа, Стьюдента.
- непрерывная на $[a, \infty)$
  - Гамма
- непрерывная на $[a, b]$
  - Бета
- дискретное на $[a, \infty)$
  - Пуассона
- дискретное на $[a, b]$
  - Биномиальная

Сэмплирование из смеси

1. Сэмплируем компоненты z с вероятностями $\phi_1, \phi_2, ...\phi_K$

Сэмплирование из смеси

1. Сэмплируем компоненты z с вероятностями $\phi_1, \phi_2, ...\phi_K$
    - сэмплируем $u \sim Uniform[0,1]$ и выбираем компоненту k, если $\sum_{i=1}^{k-1} \phi_i < u \leq \sum_{i=1}^{k} \phi_i$
2. Сэмплируем наблюдение $x \sim p(x|\theta_k)$

Классификация с использованием смесей

Байесовское правило минимальной ошибки:

$$\widehat{y} = \arg\max_y p(y|x) = \arg\max_y \frac{p(y)p(x|y)}{p(x)} = \arg\max_y p(y)p(x|y)$$

$p(y)$ - априорная вероятность класса y
$p(x|y)$ - внутриклассовое распределение x

Классификация с использованием смесей

Байесовское правило минимальной ошибки:

$$\widehat{y} = \arg\max_y p(y|x) = \arg\max_y \frac{p(y)p(x|y)}{p(x)} = \arg\max_y p(y)p(x|y)$$

$p(y)$ - априорная вероятность класса y
$p(x|y)$ - внутриклассовое распределение x

Моделируем внутриклассовое распределение смесью:

$$p(x|y) = \sum_{k=1}^{K_y} \phi_{y,k} p(x; \theta_{y,k})$$

где K_y, $\pi_{y,k}$ и $p(x;\theta_{y,k})$ свои для каждого класса y.

EM-алгоритм для оценки смеси Гауссиан

Инициализировать ϕ_k, μ_k, Σ_k для $k = 1, 2, ...K$.

ПОВТОРЯТЬ до сходимости:

Е-ШАГ. расчитать соответствия x_n каждой компоненте k:

ДЛЯ $n = 1, 2, ...N$:

ДЛЯ $k = 1, 2, ...K$:

$$w_{nk} = \frac{\widehat{\phi}_k N(x_n; \widehat{\mu_k}, \widehat{\Sigma}_k)}{\sum_i \widehat{\phi}_i N(x_n; \widehat{\mu}_i, \widehat{\Sigma}_i)}$$ # =p(k|x(n))

М-ШАГ. Обновить параметры смеси:

FOR $k = 1, 2, ...K$:

$$\widehat{\phi}_k = \frac{1}{N} \sum_{n=1}^{N} w_{nk}$$

$$\widehat{\mu}_k = \frac{\sum_{n=1}^{N} w_{nk} x_n}{\sum_{n=1}^{N} w_{nk}}$$

$$\widehat{\Sigma}_k = \frac{1}{\sum_{n=1}^{N} w_{nk}} \sum_{n=1}^{N} w_{nk} (x_n - \widehat{\mu}_k)(x_n - \widehat{\mu}_k)^T$$

Интерпретация

$$\begin{aligned} w_{nk} &= P(z_n = k|x_n) = \frac{P(k, x_n)}{P(x_n)} = \frac{P(k, x_n)}{\sum_i P(i, x_n)} = \\ &= \frac{P(k)P(x_n|k)}{\sum_i P(i)P(x_n|i)} = \frac{\widehat{\phi}_k N(x_n; \widehat{\mu}_k, \widehat{\Sigma}_k)}{\sum_i \widehat{\phi}_i N(x_n; \widehat{\mu}_i, \widehat{\Sigma}_i)} \end{aligned}$$

$\widehat{\phi}_k, \widehat{\mu}_k, \widehat{\Sigma}_k$ - взвешенные средние с учетом степени соответствия $w_{nk} = P(z_n = k|x_n)$:

$$\widehat{\phi}_k = \frac{1}{N} \sum_{n=1}^{N} w_{nk} \qquad \widehat{\mu}_k = \frac{\sum_{n=1}^{N} w_{nk} x_n}{\sum_{n=1}^{N} w_{nk}}$$

$$\widehat{\Sigma}_k = \frac{1}{\sum_{n=1}^{N} w_{nk}} \sum_{n=1}^{N} w_{nk} \left(x_n - \widehat{\mu}_k\right) \left(x_n - \widehat{\mu}_k\right)^T \tag{1}$$

Содержание

Алгоритм K-средних

- Стоит задача разбиения N объектов на K кластеров.
- Центр кластера k: $\mu_k, k = \overline{1, K}$.
- x_n ассоциируется кластеру z_n, $n = \overline{1, N}$.
- В кластеризации решается задача:

$$\sum_{n=1}^{N} \rho(x_n, \mu_{z_n})^2 \to \min_{z_1,...z_N,\mu_1,...\mu_K} \tag{2}$$

- Полная оптимизация вычислительно трудоёмка.
- K-средних оптимизирует (2) методом покоординатного спуска.
    - уточняем $z, \mu, z, \mu, ...$

Алгоритм K-средних

```
Инициализировать μ_k , k = 1, 2, ...K
# обычно, выбором случайных объектов x(n)
ПОВТОРЯТЬ до сходимости:
    ДЛЯ i = 1, 2, ...N:   # назначение кластеров
        z_i = arg min_{k∈{1,2,...K}} ||x_i − μ_k||
    ДЛЯ k = 1, 2, ...K:   # пересчет средних
        μ_k = (1/N_k) ∑_{n:z_n=k} x_n
```

Возможные условия остановки:

- достигнуто макс. # итераций
- назначения кластеров $z_1, ...z_N$ перестали меняться (точный критерий)
- центры кластеров $\{\mu_k\}_{k=\overline{1,K}}$ перестали изменяться существенно (приближенный критерий)

Кластеризация с помощью оценки смеси

- Для каждого x_n оценка смеси K компонентами

$$x_n \longrightarrow w_{nk} = p(k|x_n) \in [0, 1]$$

- Это
  - мягкая кластеризация (soft clustering) $w_{nk} \in [0, 1]$ на K кластеров
  - с априорными вероятностями $\phi_1, ...\phi_K$
  - распределениями кластеров $p(x;\theta_1), ...p(x;\theta_K)$.

Кластеризация с помощью смеси и K-средних

- EM кластеризация становится K-средних, когда

Кластеризация с помощью смеси и K-средних

- EM кластеризация становится K-средних, когда
- компоненты смеси - Гауссианы
  - с равными априорными вероятностями $\phi_1 = ... = \phi_K = \frac{1}{K}$
  - с ковариационными матрицами σI, $\sigma \to 0$

$$w_{nk} = \frac{\phi \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu_k)^2}{2\sigma^2}}}{\sum_i \phi \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma^2}}} = \frac{e^{-\frac{(x-\mu_k)^2}{2\sigma^2}}}{\sum_i e^{-\frac{(x-\mu_i)^2}{2\sigma^2}}} \xrightarrow{\sigma\to 0} [0...0, 1, 0, ...0]$$

Инициализация смеси Гауссиан

1. Применим K-средних к $x_1, x_2, ...x_N$, получим $\{\mu_k\}_k$ и $\{z_n\}_n$.
2. Инициализируем априорные вероятности компонент смеси

$$\widehat{\phi}_k = \frac{N_k}{N}$$

3. Инициализируем средние через μ_k, $k = \overline{1, K}$.
4. Инициализируем матрицы ковариации

$$\widehat{\Sigma}_k = \frac{1}{N_k} \sum_{n:z_n=k} (x_n - \mu_k)(x_n - \mu_k)^T$$

Свойства оценки смеси через EM

- Существует много локальных максимумов
  - в частности правдоподобие$\to \infty$ при $\mu_k = x_n$ и $\sigma_k \to 0$ $\forall n$.
- EM алгоритм находит отдельный локальный оптимум
  - в зависимости от инициализации
- Можно перезапустить несколько раз и выбрать решение с наибольшим правдоподобием.
- Число компонент смеси может быть выбрано:

Свойства оценки смеси через EM

- Существует много локальных максимумов
  - в частности правдоподобие$\to \infty$ при $\mu_k = x_n$ и $\sigma_k \to 0$ $\forall n$.
- EM алгоритм находит отдельный локальный оптимум
  - в зависимости от инициализации
- Можно перезапустить несколько раз и выбрать решение с наибольшим правдоподобием.
- Число компонент смеси может быть выбрано:
  - кросс-валидацией на итоговой задаче
  - вневыборочным значением правдоподобия
  - информационными критериями (l=log(правдоподобие), k =#параметров)

$$AIC = 2k - 2l$$
$$BIC = k \ln N - 2l$$

  - BIC состоятельно оценивает порядок авторегрессионной модели, AIC переоценивает.

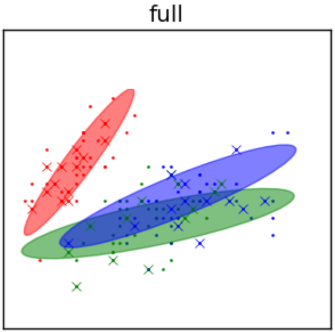
Содержание

Упрощение смеси Гауссиан

- $\Sigma_k \in \mathbb{R}^{DxD}$ требует $\frac{D(D+1)}{2}$ параметров.
- Для всех компонент: $K\frac{D(D+1)}{2}$ параметров.
- Компоненты плохо определяются, когда
    - $K\frac{D(D+1)}{2}$ велико по сравнению с N
    - компоненты лежат рядом друг с другом
- Можем улучшить оценивание упрощениями Σ_k.

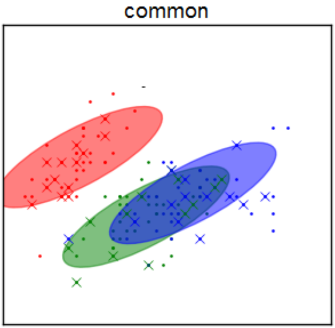
Ковариационные матрицы без ограничений

Σ_k без ограничений



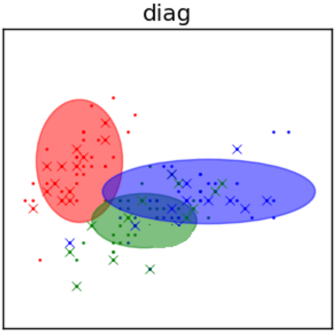
Общие ковариационные матрицы

$$\Sigma_1 = ... = \Sigma_K$$



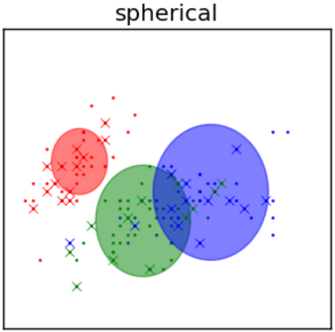
Diagonal covariance matrices

$$\Sigma_k = \text{diag}\{\sigma^2_{k,1}, \sigma^2_{k,2} ... \sigma^2_{k,D}\}$$



Сферические ковариационные матрицы

$\Sigma_k = \sigma_k^2 I$, $I \in \mathbb{R}^{DxD}$ - единичная матрица



Заключение

- Смеси распределений - обобщение оценкой одним распределением.
  - с $\uparrow K$ гибкость смеси растет
  - $p_i(x)$ - любые, не обязательно Гауссовы
- Параметры смеси находятся EM алгоритмом.
- Число компонент можно найти по вневыборочному правдоподобию.
- Смесь Гауссиан - обобщение метода K-средних.