

Смеси распределений

Виктор Китов

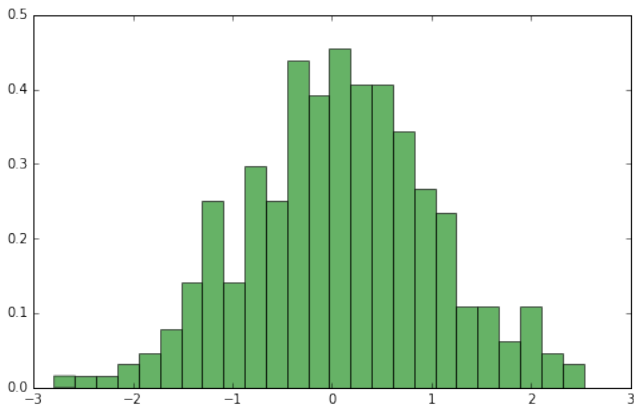
v.v.kitov@yandex.ru

Содержание

- 1 Смеси распределений
- 2 К-средних
- 3 Упрощение смеси Гауссиан

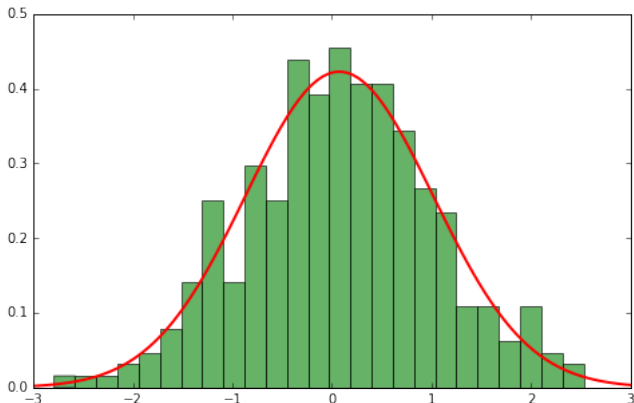
Выборочная плотность

Рассмотрим выборочную плотность:



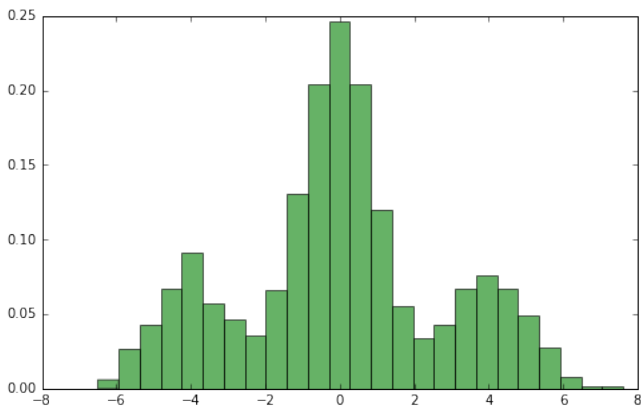
Параметрическая аппроксимация плотности

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

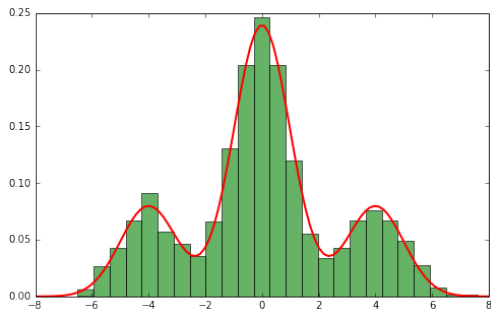
$p(x, \theta_k)$ м. быть из одинакового или разных параметрических семейств.

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 $\phi_1, \phi_2, \dots, \phi_K$

Сэмплирование из смеси

- 1 Сэмплируем компоненты z с вероятностями $\phi_1, \phi_2, \dots, \phi_K$
 - сэмплируем $u \sim \text{Uniform}[0, 1]$ 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:

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

$p(y)$ - априорная вероятность класса y

$p(x|y)$ - внутриклассовое распределение x

Классификация с использованием смесей

Байесовское правило минимальной ошибки:

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

$p(y)$ - 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 .

ЕМ-алгоритм для оценки смеси Гауссиан

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

ПОВТОРЯТЬ до сходимости:

Е-ШАГ. рассчитать соответствия x_n
каждой компоненте k :

ДЛЯ $n = 1, 2, \dots, N$:

ДЛЯ $k = 1, 2, \dots, K$:

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

М-ШАГ. Обновить параметры смеси:

FOR $k = 1, 2, \dots, K$:

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

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

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

Интерпретация

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

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

$$\begin{aligned}
 \hat{\phi}_k &= \frac{1}{N} \sum_{n=1}^N w_{nk} & \hat{\mu}_k &= \frac{\sum_{n=1}^N w_{nk} x_n}{\sum_{n=1}^N w_{nk}} \\
 \hat{\Sigma}_k &= \frac{1}{\sum_{n=1}^N w_{nk}} \sum_{n=1}^N w_{nk} (x_n - \hat{\mu}_k)(x_n - \hat{\mu}_k)^T
 \end{aligned} \tag{1}$$

Содержание

- 1 Смеси распределений
- 2 К-средних**
- 3 Упрощение смеси Гауссиан

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:

$$\sum_{n=1}^N \rho(x_n, \mu_{z_n})^2 \rightarrow \min_{z_1, \dots, z_N, \mu_1, \dots, \mu_K} \quad (2)$$

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

Алгоритм К-средних

Инициализировать μ_k , $k = 1, 2, \dots, K$

обычно, выбором случайных объектов $x(n)$

ПОВТОРЯТЬ до сходимости:

ДЛЯ $i = 1, 2, \dots, N$: # назначение кластеров

$$z_i = \arg \min_{k \in \{1, 2, \dots, K\}} \|x_i - \mu_k\|$$

ДЛЯ $k = 1, 2, \dots, K$: # пересчет средних

$$\mu_k = \frac{1}{N_k} \sum_{n: z_n = k} x_n$$

Возможные условия остановки:

- достигнуто макс. # итераций
- назначения кластеров z_1, \dots, z_N перестали меняться (точный критерий)
- центры кластеров $\{\mu_k\}_{k=1, \overline{K}}$ 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 ϕ_1, \dots, ϕ_K
 - распределениями кластеров $p(x; \theta_1), \dots, p(x; \theta_K)$.

Кластеризация с помощью смеси и К-средних

- EM кластеризация становится К-средних, когда

Кластеризация с помощью смеси и К-средних

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

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

Инициализация смеси Гауссиан

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

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

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

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

Свойства оценки смеси через EM

- Существует много локальных максимумов
 - в частности правдоподобие $\rightarrow \infty$ при $\mu_k = x_n$ и $\sigma_k \rightarrow 0 \forall n$.
- 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 $\rightarrow \infty$ при $\mu_k = x_n$ и $\sigma_k \rightarrow 0 \forall n$.
- 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
($l = \log(\text{правдоподобие})$, $k = \# \text{параметров}$)

$$AIC = 2k - 2l$$

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

- BIC состоятельно оценивает порядок авторегрессионной модели, AIC переоценивает.

Содержание

- 1 Смеси распределений
- 2 К-средних
- 3 Упрощение смеси Гауссиан

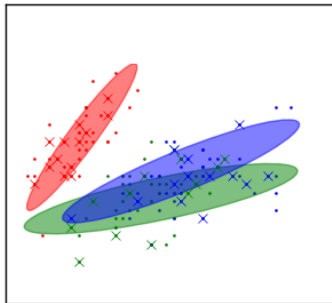
Упрощение смеси Гауссиан

- $\Sigma_k \in \mathbb{R}^{D \times D}$ 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

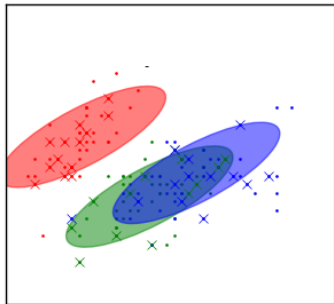
full



Общие ковариационные матрицы

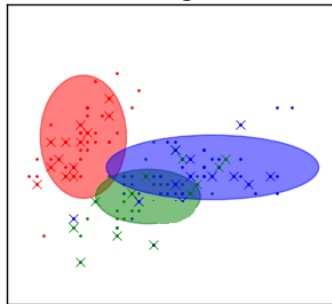
$$\Sigma_1 = \dots = \Sigma_K$$

common



Diagonal covariance matrices

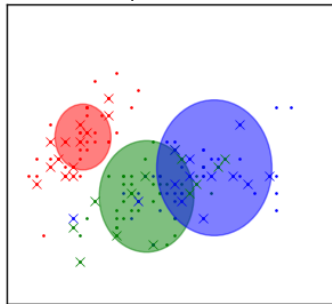
$$\Sigma_k = \text{diag}\{\sigma_{k,1}^2, \sigma_{k,2}^2 \dots \sigma_{k,D}^2\}$$



Сферические ковариационные матрицы

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

spherical



Заключение

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