

ЗАДАЧА РОЗДІЛУ СУМІШІ.

**АЛГОРИТМ EXPECTATION-
MAXIMIZATION (EM)**

Постановка задачі розділу суміші

У випадках коли форму класу не вдається описати одним розподілом, намагаються описати її сумішшю розподілів.

“Вибрати об’єкт x із суміші $p(x)$ ” означає спочатку вибрати j -ту компоненту суміші з дискретного розподілу $\{p(j) \mid j = 1, \dots, k\}$, а потім вибрати об’єкт x відповідно до щільності $p(x \mid j) = p_j(x)$.

Щільність розподілу на множині X як суміш k розподілів:

$$p(x) = \sum_{j=1}^k p(j)p(x \mid j) = \sum_{j=1}^k w_j p_j(x) \quad \sum_{j=1}^k w_j = 1 \quad w_j \geq 0$$

$p(j) = w_j$ - апіорна імовірність j -ї компоненти суміші,
ваговий коефіцієнт в сумі,

$p_j(x)$ - щільність розподілу або функція правдоподібності j -ї компоненти суміші.

Нехай функції щільності відрізняються лише значеннями параметрів:

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

Базовий алгоритм Expectation-Maximization (EM)

Дано: вибірка X з m випадкових і незалежних спостережень x із суміші $p(x)$, відоме число k . Відома функція φ .

Необхідно: оцінити вектор параметрів $\Theta = (w_1, \dots, w_k, \theta_1, \dots, \theta_k)$.

Загальна ідея алгоритму EM:

- 1) Обчислити початкове наближення вектору параметрів Θ
- 2) повторювати

$$Q := EStep(\Theta)$$

$$\Theta := MStep(\Theta, Q)$$

- 3) поки Q і/або Θ не стабілізуються

Е-крок базового алгоритму ЕМ

Е-крок: розраховується очікуване значення вектору скритих змінних Q на основі поточного наближення вектора параметрів Θ

Нехай $p(x; \theta_j)$ - щільність імовірності того, що об'єкт x отримано з j -ї компоненти суміші:

$$p(x; \theta_j) = p(x)p(\theta_j | x) = w_j p_j(x)$$

Позначимо $q_{ij} = p(\theta_j | x_i)$ - невідома апостеріорна імовірність того, що об'єкт x_i отримано з j -ї компоненти суміші.

Візьмемо q_{ij} в якості **скритих змінних**.

$$\sum_{j=1}^k q_{ij} = 1 \quad \forall i = 1, \dots, m \quad q_{ij} = \frac{w_j p_j(x_i)}{\sum_{s=1}^k w_s p_s(x_i)} \quad \forall i, j$$

М-крок базового алгоритму EM

М-крок: розв'язується задача максимізації правдоподібності і знаходиться наступне наближення вектору Θ на основі поточних значень векторів Q і Θ :

$$R(\Theta) = \ln \prod_{i=1}^m p(x_i) = \sum_{i=1}^m \ln \sum_{j=1}^k w_j p_j(x_i) \rightarrow \max_{\Theta}$$

при обмеженні $\sum_{j=1}^k w_j = 1$

Розв'язок цієї задачі оптимізації: $w_j = \frac{1}{m} \sum_{i=1}^m q_{ij} \quad j = 1, \dots, k$

$$\theta_j = \arg \max_{\theta} \sum_{i=1}^m q_{ij} \ln \varphi(x_i, \theta)$$

Отримано k незалежних оптимізаційних задач.

Розділ змінних має місце внаслідок вдалого вибору скритих змінних.

Обґрунтування М-кроку, виведення формул

Розв'язання задачі методом множників Лагранжа

$$L(\Theta, X) = \sum_{i=1}^m \ln \sum_{j=1}^k w_j p_j(x_i) - \lambda \left(\sum_{j=1}^k w_j - 1 \right)$$

$$\frac{\partial L}{\partial w_j} = \sum_{i=1}^m \frac{p_j(x_i)}{\sum_{s=1}^k w_s p_s(x_i)} - \lambda = 0 \quad j = 1, \dots, k \quad (*)$$

Помножимо ліву і праву частини (*) на w_j , візьмемо суму всіх k рівностей :

$$\sum_{i=1}^m \sum_{j=1}^k \frac{w_j p_j(x_i)}{\sum_{s=1}^k w_s p_s(x_i)} = \lambda \sum_{j=1}^k w_j \Rightarrow \lambda = m$$

Помножимо ліву і праву частини (*) на w_j , підставимо $\lambda = m$:

$$w_j = \frac{1}{m} \sum_{i=1}^m \frac{w_j p_j(x_i)}{\sum_{s=1}^k w_s p_s(x_i)} = \frac{1}{m} \sum_{i=1}^m q_{ij}$$

Обґрунтування М-кроку, виведення формул

Розв'язання задачі методом множників Лагранжа

$$L(\Theta, X) = \sum_{i=1}^m \ln \sum_{j=1}^k w_j p_j(x_i) - \lambda \left(\sum_{j=1}^k w_j - 1 \right)$$

$$\begin{aligned} \frac{\partial L}{\partial \theta_j} &= \sum_{i=1}^m \frac{w_j}{\sum_{s=1}^k w_s p_s(x_i)} \frac{\partial}{\partial \theta_j} p_j(x_i) = \sum_{i=1}^m \frac{w_j p_j(x_i)}{\sum_{s=1}^k w_s p_s(x_i)} \frac{\partial}{\partial \theta_j} \ln p_j(x_i) \\ & \qquad \qquad \qquad j = 1, \dots, k \\ &= \sum_{i=1}^m q_{ij} \frac{\partial}{\partial \theta_j} \ln p_j(x_i) = \frac{\partial}{\partial \theta_j} \sum_{i=1}^m q_{ij} \ln p_j(x_i) = 0 \qquad p_j(x) = \varphi(x; \theta_j) \end{aligned}$$

Отримана умова співпадає з необхідною умовою максимуму в задачі максимізації зваженої правдоподібності:

$$\theta_j = \arg \max_{\theta} \sum_{i=1}^m q_{ij} \ln \varphi(x_i, \theta) \qquad j = 1, \dots, k$$

Алгоритм ЕМ з фіксованою кількістю компонент

Дано:

вибірка даних $X = (x_1, \dots, x_m)$

k – число компонентів суміші

$\Theta = ((w_j, \theta_j) \mid j = 1, \dots, k)$ – початкове наближення параметрів суміші

δ – параметр критерію зупинки алгоритму

Знайти:

$\Theta = ((w_j, \theta_j) \mid j = 1, \dots, k)$ – оптимізований вектор параметрів суміші

Алгоритм ЕМ з фіксованою кількістю компонент

Повторювати

Е-крок:

$$\text{для всіх } \begin{matrix} i = 1, \dots, m \\ j = 1, \dots, k \end{matrix} \quad q_{ij}^0 := q_{ij} \quad q_{ij} := \frac{w_j \varphi(x_i, \theta_j)}{\sum_{s=1}^k w_s \varphi(x_i, \theta_s)}$$

М-крок:

$$\text{для всіх } j = 1, \dots, k$$
$$w_j = \frac{1}{m} \sum_{i=1}^m q_{ij} \quad \theta_j = \arg \max_{\theta} \sum_{i=1}^m q_{ij} \ln \varphi(x_i, \theta)$$

Поки $\max_{i,j} |q_{ij}^0 - q_{ij}| > \delta$

Сума береться за всіма елементами навчальної вибірки з ваговими коефіцієнтами q_{ij} .

Вивести вектор $\Theta = ((w_j, \theta_j) \mid j = 1, \dots, k)$

Недоліки базового алгоритму EM

- ❑ алгоритм нестійкий до вибору початкового значення вектора параметрів, оскільки він знаходить локальний максимум, бо оптимізаційна функція $R(\Theta)$ невивукла;
- ❑ швидкість збіжності може бути повільною;
- ❑ необхідно знати k – кількість компонентів суміші
 - візуально оцінити, спроектувавши вибірку на площину,
 - залучити експертів,
 - розв'язувати задачу декілька разів при різних значеннях k , побудувати графік залежності $R(\Theta)$ від k , вибрати найменше k , при якому має місце різкий скачок $R(\Theta)$ (критерій “крутого схилу”).

Узагальнений алгоритм ЕМ

Розв'язуючи задачу оптимізації

$$\theta_j = \arg \max_{\theta} \sum_{i=1}^m q_{ij} \ln \varphi(x_i, \theta) ,$$

достатньо зробити одну або декілька ітерацій в напрямку знаходження максимуму і перейти на Е-крок.

Цей алгоритм також має непогану збіжність.

Стохастичний алгоритм ЕМ

- ❑ Результати стохастичного алгоритму практично не залежать від початкового наближення вектора параметрів.
- ❑ Стохастичний алгоритм знаходить екстремум, близький до глобального.

Ідея: випадковим чином, але ціленаправлено, “встряхувати” навчальну вибірку на кожній ітерації.

Це “вибиває” оптимізаційний процес з локальних максимумів.

Стохастичний алгоритм EM

Нехай навчальна вибірка X розбита на компоненти суміші K_j ,
кожний об'єкт x_i відноситься до єдиної K_j , $j = 1, \dots, k$.

Ітераційно повторюються три кроки:

- 1) S-крок: стохастичне моделювання
- 2) E-крок: без змін
- 3) M-крок: розв'язується задача максимізації *незв'язаної* правдоподібності, і знаходиться наступне наближення вектору Θ на основі поточних значень векторів Q і Θ .

Стохастичний алгоритм EM

S-крок: стохастичне моделювання

$\forall i = 1, \dots, m$ генерується вектор $y_i = (y_{i1}, y_{i2}, \dots, y_{ik})$, який має поліноміальний розподіл з параметрами 1 і q_{ij} : $p(y_{ij} = 1) = q_{ij}$.

За векторами y_i визначається розбиття вибірки X на K_j , а також кількість елементів у кожному K_j : v_1, v_2, \dots, v_k .

М-крок:
$$w_j = \frac{v_j}{m} \quad j = 1, \dots, k$$

Розв'язується задача максимізації незваженої правдоподібності

$$\theta_j = \arg \max_{\theta} \sum_{x_i \in K_j} \ln \varphi(x_i, \theta) \quad j = 1, \dots, k$$

Алгоритм ЕМ з послідовним додаванням компонент (розв'язання задачі кластеризації)

Дано:

вибірка даних $X = (x_1, \dots, x_m)$

R – максимально допустимий розкид правдоподібності об'єктів,

m_0 – мінімальна довжина вибірки, за якою можна відновити щільність,

δ – параметр критерію зупинки.

Знайти:

k – число компонентів суміші,

$\Theta = ((w_j, \theta_j) \mid j = 1, \dots, k)$ – вектор параметрів суміші.

Цей алгоритм дозволяє розв'язати як проблему вибору кількості компонент, так і проблему вибору початкового наближення.

Алгоритм ЕМ з послідовним додаванням компонент (розв'язання задачі кластеризації)

1) початкове наближення – одна компонента

$$\theta_1 := \arg \max_{\theta} \sum_{i=1}^m \ln \varphi(x_i, \theta) \quad w_1 := 1 \quad k := 1$$

2) для всіх $k = 2, 3, \dots$

- знайти об'єкти з низьким значенням правдоподібності:

$$U := \{x_i \in X : p(x_i) < \max_j p(x_j) / R\}$$

- якщо $|U| < m_0$, то вихід з циклу по k

- початкове наближення для k -ої компоненти:

$$\theta_k := \arg \max_{\theta} \sum_{x_i \in U} \ln \varphi(x_i, \theta) \quad w_k := |U| / m$$

$$w_j := w_j (1 - w_k) \quad j = 1, \dots, k - 1$$

- виконати базовий алгоритм ЕМ(X, k, Θ, δ)