Mezcla de Gaussianas

Mezcla de Gaussianas

Tags: Reconocimiento de Patrones, Métodos de maxima verosimilitud

---

Para conjuntos de datos complejos (con relaciones no aparentes), es normal modelar su distribución de probabilidad con múltiples funciones gaussianas en lugar de una sola.

Suponer que se tiene un conjunto de N datos independientes $x = x_1, x_2, \ldots, x_N$. Se puede modelar este conjunto de datos con un modelo de mezcla de Gaussianas.

Suponer que se tiene k distribuciones gaussianas

$$N_k(x_i, \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} e^{-1/2(x_i-\mu_k)^T \Sigma_k^{-1}(x_i-\mu_k)}$$

Se puede etiquetar el dato x_i a una de las k distribuciones si

$$b_{ik} = \begin{cases} 1, \text{si } x_i \text{ pertenece a la distribución } N_k \\ 0, \text{si no} \end{cases}$$

Definiendo $a_k = P(b_{ik} = 1)$

Donde se cumple que:

$$\sum_{i=1}^{K} a_i = 1$$

donde K es el número de clases (Gaussianas)

$$\sum_{i=1}^{N} \sum_{k=1}^{K} b_{ik} = N = \text{ número de datos en el conjunto}$$

Por otro lado, tenemos la probabilidad de que el dato x_i pertenezca a la distribución N_k dado b_{ik}, μ_k y Σ_k, esto es:

$$P(x_i | b_{ik} = 1, \mu_k, \Sigma_k) = N_k(x_i, \mu_k, \Sigma_k)$$

donde μ_k son las medias de las k gaussianas y Σ_k las matrices de covarianza.

Para calcular la verosimilitud, se asume independencia en los datos.

$$P(A \cap B) = P(A)P(B) \qquad (1.3.15)$$

$$P(w_j|x) = \frac{P(x|w_j)P(w_j)}{P(x)} \rightarrow \text{Regla de Bayes} \qquad (2.1.5)$$

Teorema de Bayes

$$P(A_j|E) = \frac{P(A_j)P(E|A_j)}{\sum_{i=1}^{k} P(A_j)P(E|A_i)}, j = 1, 2, \ldots, k \qquad (1.3.14)$$

$$l(\theta) = \prod_{i=1}^{N}\prod_{k=1}^{K}[P(x_i|b_{ik}, \mu_k, \Sigma_k)P(b_{ik} = 1)]^{b_{ik}}$$

Si se calcula la log-similitud

$$\alpha(\theta) = \log\,(l(\theta)),\ \text{entonces}$$

$$\alpha(\theta) = \sum_{i=1}^{N}\sum_{k=1}^{K} b_{ik}[\log P(x_i)|b_{ik}, \mu_k, \Sigma_k + \log P(b_{ik} = 1)]$$

$$\log[(a \cdot b)^2] = 2\log(a \cdot b) = 2[\log(a) + \log(b)] \qquad (1)$$

Por otro lado, considerando $\log P(x_i|b_{ik}, \mu_k, \Sigma_k)$

$$\log[P(x_i|b_{ik}, \mu_k, \Sigma_k)] = \log[\frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} e^{-1/2}(x_i - \mu_k)^T \Sigma_k^{-1}(x_i - \mu_k)]$$

Aplicando las propiedades de los logaritmos

$$\log[P(x_i|b_{ik}, \mu_k, \Sigma_k)] = \frac{-d}{2}\log(2\pi) - \frac{1}{2}\log(\Sigma_k) - \frac{1}{2}(x_i - \mu_k)^T \Sigma_k^{-1}(x_i - \mu_k) \qquad (2)$$

Sustituyendo (2) en (1):

$$\alpha(\theta) = \sum_{i=1}^{N}\sum_{k=1}^{K} b_{ik}[\frac{-d}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(x_i - \mu_k^T \Sigma_k^{-1}(x_i - \mu_k + \log(a_k))]$$

Definiendo $a_k = P(b_{ik} = 1)$

Parámetros a optimizar por medio del algoritmo EM (Expectation-Maximization)

$$\theta = \{b_{ik}, \mu_k, \Sigma_k, a_k\}$$

Para resolver se utiliza el algoritmo EM. Este algoritmo tiene 2 pasos, un paso M de maximización y un paso E del cálculo de la esperanza.

- Paso M

Se calcula el máximo de $\alpha(\theta)$ suponiendo que b_{ik} está dado por:

$$b_{ik} = \frac{1}{M_k} \curlyvee K \text{ puede ser cualquiera}$$

donde M_k es el número de datos pertenecientes a cada clase

$$\frac{\partial \alpha(\theta)}{\partial_{\mu_k}} = \frac{\partial}{\partial_{\mu_k}} \left[\sum_{i=1}^{N} \sum_{k=1}^{K} b_{ik} [-\frac{1}{2}(x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k)] \right] = 0$$

$$\frac{\partial \alpha(\theta)}{\partial_{\mu_k}} = \frac{\partial}{\partial_{\mu_k}} \left[\sum_{i=1}^{N} \sum_{k=1}^{K} b_{ik} (-\frac{1}{2})(x_i^2 - 2x_i\mu_k + \mu_k^2)\Sigma_k^{-1} \right] = 0$$

$$\frac{\partial \alpha(\theta)}{\partial_{\mu_k}} = \sum_{i=1}^{N} [b_{ik}x_i - b_{ik}\mu_k] = 0$$

$$\sum_{i=1}^{N} b_{ik}x_i - \sum_{i=1}^{N} b_{ik}\mu_k = 0$$

$$\mu_k = \frac{\sum_{i=1}^{N} b_{ik}x_1}{\sum_{i=1}^{N} b_{ik}}$$

Para calcular la matriz de covarianza Σ_k:

$$\frac{\partial \alpha(\theta)}{\partial_{\Sigma_k}} = 0$$

$$\frac{\partial \alpha(\theta)}{\partial_{\Sigma_k}} = \sum_{i=1}^{N} \frac{\partial}{\partial_{\Sigma_k}} [b_{ik}(-\frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(x_i - \mu_k)^T \Sigma_k^{-1}(x_i - \mu_k))]$$

$$\frac{\partial \alpha(\theta)}{\partial_{\Sigma_k}} = \sum_{i=1}^{N} (-b_{ik} + \Sigma_k^{-1} b_{ik}(x_i - \mu_k)^T (x_i - \mu_k)) = 0$$

$$\Sigma_k = \frac{\sum_{i=1}^{N} b_{ik}(x_i - \mu_k)^T (x_i - \mu_k)}{\sum_{i=1}^{N} b_{ik}}$$

Utilizando multiplicadores de Lagrange

$$\mathcal{L}(x, \lambda) = f(x) - \lambda g(x)$$

Donde:

$$\begin{aligned} &\mathcal{L} \rightarrow \text{Lagrangiano} \\ &\lambda \rightarrow \text{Multiplicador} \\ &g(x) \rightarrow \text{Limitación de igualdad} \\ &f(x) \rightarrow \text{Función} \end{aligned}$$

$$\frac{\partial \alpha(\theta)}{\partial_{a_k}} = \frac{\partial}{\partial_{a_k}} \left[\sum_{i=1}^{N} \sum_{k=1}^{K} b_{ik} \log(ak) \right] = 0$$

limitación

$$\sum_{i=1}^{K} a_i = 1 \quad \therefore \quad \sum_{i=1}^{K} a_k - 1 = 0$$

$$\frac{\partial \alpha(\theta)}{\partial_{a_k}} = \sum_{i=1}^{N} \frac{\partial}{\partial_{a_k}} [b_{ik} \log(a_k) - \lambda(a_k - 1)] = 0$$

$$b_{ik}(\frac{1}{a_k}) - \lambda = 0$$

$$b_{ik} = a_k \lambda \text{ pero } \sum_{i=1}^{N} \sum_{k=1}^{K} b_{ik} = \lambda \sum_{k=1}^{K} a_k$$

$$b_{ik} = \frac{N_k(x_i, \mu_k, \Sigma_k) a_k}{\sum_{k=1}^{K} N_k(x_i, \mu_k, \Sigma_k) \alpha_k}$$

$$a_k = \frac{\sum_{i=1}^{N} b_{ik}}{N}$$

References

Caso multicategoria

Densidad multivariada