

clustering

Nhung Đào Thị Hồng

March 2022

1 Kmean

$$J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} ||x_n - \mu_k||^2$$

+ Step 1: Chọn các centroid tùy ý

+ Step 2: Fix μ_k , tìm r_{nk}

$$L = \sum_{k=1}^K r_{nk} ||x_n - \mu_k||^2$$

Chọn r_{ij} dựa vào khoảng cách từ điểm này tới centroid nhỏ nhất.

+ Step 3: Fix r_{nk} , find μ_k

$$J = \sum_{k=1}^K \sum_{n=1}^N r_{nk} ||x_n - \mu_k||^2$$

Vì r_{nk} đã biết nên ta biết chỉ có 1 r là có giá trị bằng 1

J chuyển về:

$$J = \sum_{n=1}^N r_{nk} ||x_n - \mu_k||^2$$

$$\frac{\sigma J}{\sigma \mu_k} = - \sum_{n=1}^N 2 \cdot r_{nk} (x_n - \mu_k) = 0$$

$$\mu_k = \frac{\sum_{n=1}^N r_{nk} \mu_k}{\sum_{n=1}^K r_{nk}}$$

-> μ_k là trung bình cộng trong clustering thứ k

2 Gaussian mixture model

3 BDscan