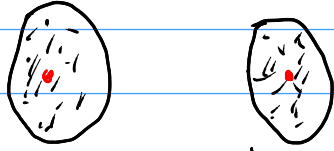


K-Means Clustering + Mixture Models

Unsupervised Learning

Clustering



$$n=5, k=3$$

$$C_1 = \{x_1, x_5\}$$

$$C_2 = \{x_2\}$$

$$C_3 = \{x_3, x_4\}$$

Given $X = \{x_1, \dots, x_n\}$, partition in C_1, \dots, C_k .

Goals (informally):

- points in a cluster are similar
- points in diff clusters are dissimilar

hyperparameter

Minimize $\sum_{j=1}^k W(C_j)$
Partition C_1, \dots, C_k
Cost fn for points in cluster j

K-means:

$$W(C_j) = \frac{1}{|C_j|} \sum_{\substack{x_i, x_{i'} \\ \in C_j}} \|x_i - x_{i'}\|_2^2 = 2 \sum_{x_i \in C_j} \|x_i - \mu_j\|_2^2$$
$$= \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i$$

Slow Alg: Try all partitions: k^n partitions

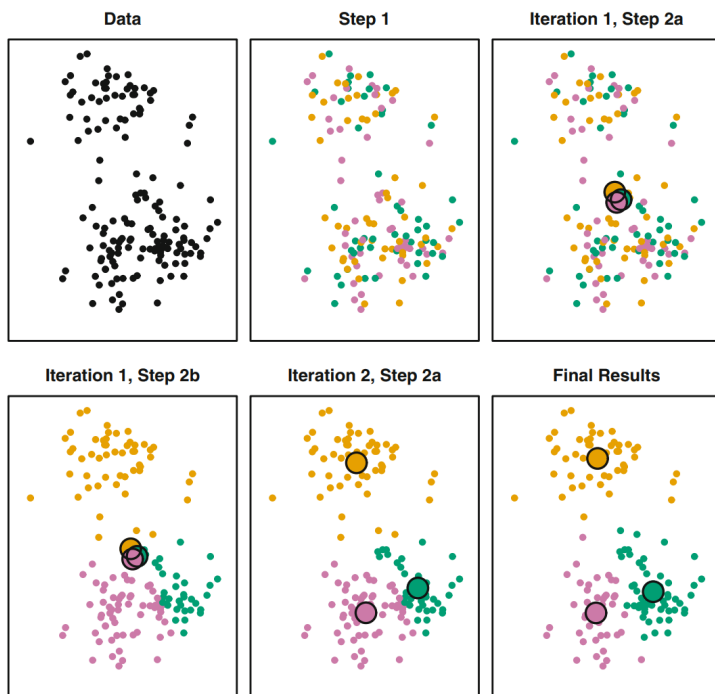
Alg:

- Initialize partition C_1, \dots, C_k

- For each cluster C_j , compute centroid $\mu_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i$

- For each point x_i , assign it to cluster w/ nearest centroid: $\operatorname{argmin}_j \|x_i - \mu_j\|_2$.

Repeat until converged



Drawbacks

- Slow converge
- Local optimum

Solutions:

- Repeat many times
- Better initialization (K-means++)



Generative Models

X_1, \dots, X_n , drawn i.i.d. from some dist p_θ

Goal: Output $\hat{p} \approx p_\theta$

Simple: $X_1, \dots, X_n \sim N(\mu, 1)$ $p_\mu(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2}\right)$

Maximum likelihood est (MLE):

$$\hat{\mu} = \underset{\mu}{\operatorname{argmax}} \sum_{i=1}^n \log p_\mu(x_i) = \underset{\mu}{\operatorname{argmax}} \sum_{i=1}^n -(x_i - \mu)^2 \\ = \frac{1}{n} \sum x_i$$

Output $\hat{p} = N(\hat{\mu}, 1)$

Mixture Model

$$p_\theta(x) = \sum_{j=1}^k \pi_j p_{\theta^{(j)}}^{(j)}(x)$$

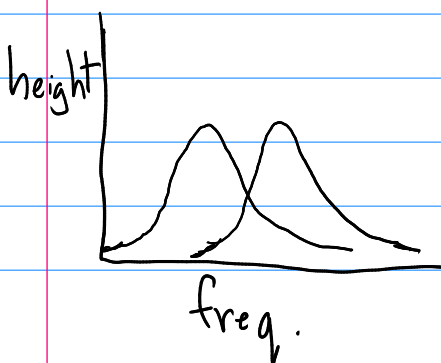
π_j 's
 $\theta^{(j)}$'s

Prob dist for component j

Mixing weight
 $\pi_j \geq 0, \sum \pi_j = 1$

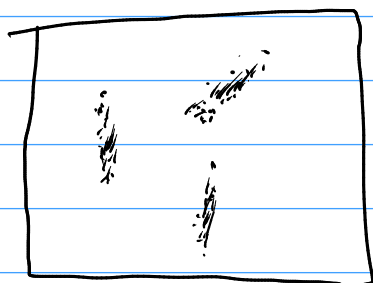
Sampling procedure:

1. Draw sample from $\{1, \dots, k\}$, from π
2. Output sample from $p_{\theta^{(j)}}^{(j)}(x)$



Gaussian Mixture Model (GMM)

$$P_{\theta}(x) = \sum_{j=1}^k \pi_j N(\mu_j, \Sigma_j, x)$$



Z_i = component that X_i was sampled from
 $Z_i \in \{1, \dots, k\}$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^n \log P_{\theta}(X_i)$$

$$= \underset{\theta = (\mu_j, \Sigma_j, \pi_j)}{\operatorname{argmax}} \sum \log \left(\sum_{j=1}^k \mathbb{1}\{z_i=j\} \pi_j N(\mu_j, \Sigma_j, X_i) \right)$$

- Use $\hat{\pi}_j = \frac{1}{n} \sum_{z_i=j} 1$

$$\hat{\mu}_j = \frac{1}{\sum_{z_i=j} 1} \sum_{z_i=j} X_i, \quad \hat{\Sigma}_j = \frac{1}{\sum_{z_i=j} 1} \sum_{z_i=j} (X_i - \hat{\mu}_j)(X_i - \hat{\mu}_j)^T$$

Expectation Maximization (EM):

- "soft" version of k-means

1. Given θ , fractionally assign points X_i to clusters
2. Given fractional assignment of X_i to clusters, compute best θ .

$$\operatorname{argmax}_{\theta} \sum_{i=1}^n \log P_{\theta}(x_i)$$

$$\log P_{\theta}(x_i)$$

$$= \log \sum_{j=1}^k P_{\theta}(x_i, z_i=j)$$

$$= \log \sum_{j=1}^k \frac{q_i(z_i=j)}{q_i(z_i=j)} P_{\theta}(x_i, z_i=j)$$

$$= \log \sum_{j=1}^k q_i(z_i=j) \left(\frac{P_{\theta}(x_i, z_i=j)}{q_i(z_i=j)} \right)$$

$$= \log \mathbb{E}_{z_i \sim q_i} \left[\frac{P_{\theta}(x_i, z_i)}{q_i(z_i)} \right]$$

$$\geq \mathbb{E}_{z_i \sim q_i} \left[\log \left(\frac{P_{\theta}(x_i, z_i)}{q_i(z_i)} \right) \right]$$

$$= \sum_{j=1}^k q_i(z_i=j) \log P_{\theta}(x_i, z_i=j) - \sum_{j=1}^k q_i(z_i=j) \log q_i(z_i=j)$$

$$\operatorname{argmax}_{\theta} \sum_{i=1}^n \log P_{\theta}(x_i) \geq$$

$$\operatorname{argmax}_{\theta, q_i} \sum_{i=1}^n \sum_{j=1}^k q_i(z_i=j) \log P_{\theta}(x_i, z_i=j) - \sum_{j=1}^k q_i(z_i=j) \log q_i(z_i=j)$$

$$\operatorname{argmax}_{\theta, q_i} \sum_{i=1}^n \sum_{j=1}^K q_i(z_i=j) \log P_{\theta}(x_i, z_i=j) - \sum_{j=1}^K q_i(z_i=j) \log q_i(z_i=j)$$

E step: Fix θ , opt q_i 's

$$\operatorname{argmax}_{\text{dist } q_i} \sum_{j=1}^K q_i(z_i=j) \log P_{\theta}(x_i, z_i=j) - \sum_{j=1}^K q_i(z_i=j) \log q_i(z_i=j)$$

$$\operatorname{argmax}_{\text{dist } q_i} \left\{ \sum_{j=1}^K q_i(z_i=j) \log P_{\theta}(z_i=j | x_i) - \sum_{j=1}^K q_i(z_i=j) \log q_i(z_i=j) + \sum_{j=1}^K q_i(z_i=j) \log P_{\theta}(x_i) \right\}$$

$$= \operatorname{argmin}_{\text{dist } q_i} \sum_{j=1}^K q_i(z_i=j) \log \left(\frac{q_i(z_i=j)}{P_{\theta}(z_i=j | x_i)} \right)$$

$$KL(q_i(z_i) \| P_{\theta}(z_i | x_i))$$

Measure of dist

$$\geq 0$$

$$= 0 \text{ when } q_i(z_i) = P_{\theta}(z_i | x_i)$$

$$= P_{\theta}(z_i | x_i)$$

M Step: Fix q_i 's, opt θ

$$\operatorname{argmax}_{\theta} \sum_{i=1}^n \sum_{j=1}^K q_i(z_i=j) \log P_{\theta}(x_i, z_i=j)$$

Often solvable in closed form

Algo:

- Initialize params θ
 - Run E step
 - Run M step
- } Repeat

GMMs

$$E \text{ step: } q_i(z_i=j) = P_\theta(z_i=j | X_i) = \frac{P_\theta(z_i=j, X_i)}{P_\theta(X_i)}$$

$$= \frac{\pi_j N(\mu_j, \Sigma_j, X_i)}{\sum_{l=1}^k \pi_l N(\mu_l, \Sigma_l, X_i)} \quad (\text{compute for all } X_i, j \in \{1, \dots, k\})$$

M step: (simplify, $1D, \text{Var}=1$. $P_\theta(x) = \sum \pi_j N(\mu_j, 1, x)$)

$$\arg \max_{\theta} \sum_{i=1}^n \sum_{j=1}^k q_i(z_i=j) \log P_\theta(x_i, z_i=j)$$

$$= \arg \max_{\theta} \sum_{i=1}^n \sum_{j=1}^k q_i(z_i=j) \log \left(\pi_j \exp\left(-\frac{(x_i - \mu_j)^2}{2}\right) \right)$$

$$= \arg \max_{\theta} \sum \sum q_i(z_i=j) \left(\log \pi_j - \frac{(x_i - \mu_j)^2}{2} \right)$$

Focus on μ_j 's

$$\arg \min_{\mu_j} \sum \sum q_i(z_i=j) \frac{1}{2} (x_i - \mu_j)^2 \Rightarrow \sum_{i=1}^n -q_i(z_i=j) (x_i - \mu_j) = 0$$

$$\mu_j = \frac{\sum_{i=1}^n q_i(z_i=j) x_i}{\sum_{i=1}^n q_i(z_i=j)}$$

$$\arg \max_{\text{dist } \pi} \sum_{i=1}^n \sum_{j=1}^k q_i(z_i=j) \log \pi_j \Leftrightarrow \arg \max_{\pi} \sum_{i=1}^n \sum_{j=1}^k q_i(z_i=j) \log \pi_j + \lambda (\sum \pi_j - 1)$$

s.t. $\sum_{j=1}^k \pi_j = 1$

$$\operatorname{argmax}_{\pi} \sum_{i=1}^n \sum_{j=1}^k q_i(z_i=j) \cdot (\log \pi_j + \lambda(\sum \pi_i - 1))$$

$$\frac{d}{d\pi_j} \hookrightarrow \sum_{i=1}^n \frac{q_i(z_i=j)}{\pi_j} + \lambda = 0$$

$$\pi_j = -\frac{1}{\lambda} \sum_{i=1}^n q_i(z_i=j)$$

$$\hookrightarrow \sum_{j=1}^k \pi_j = -\frac{1}{\lambda} \sum_{i=1}^n \underbrace{\sum_{j=1}^k q_i(z_i=j)}_1 = -\frac{n}{\lambda} = 1 \Rightarrow \lambda = -n$$

$$\pi_j = \frac{1}{n} \sum_{i=1}^n q_i(z_i=j)$$

