Mixture Models

Abstract

Data $X = \{x_1, x_2, \ldots x_N\}$ $\quad x_i \in \mathbb{R}^d$

k "components"

Generative Story

Constraints $\sum_k \gamma_{ik} = 1$ and $0 \leq \gamma_{ik} \leq 1$

$$z_i \sim \text{Categorical}(\gamma_i)$$

$$x_i | z_i \sim P_\theta(z_i)$$

This obviously depends on distribution!

Model

$$P(x|\theta, \gamma) = \prod_{i=1}^{N} \sum_k \gamma_k P(x_i|\theta)$$

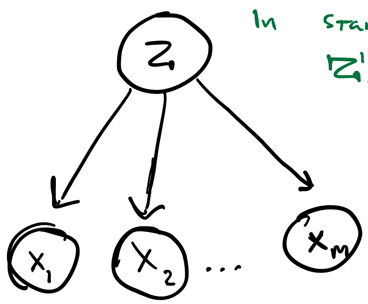
$$\lg P(x|\theta, \gamma) = \sum_{i=1}^{N} \lg\left(\sum_k \gamma_k P(x_i|\theta_k)\right)$$

Gaussian Mixture Model

$$\theta = (\mu_k, \Sigma_k)$$

$$x_i | z_i \sim \text{Gaussian}(\mu_{z_i}, \Sigma_{z_i})$$

$$\log P(X \mid \theta, \gamma) = \underbrace{\sum_{i=1}^{N} \log \Big(\sum_{k} \gamma_k \, \mathcal{N}(x_i \mid \mu_k, \Sigma_k) \Big)}_{\text{log likelihood}}$$



In standard supervised learning, Z's are observed.

If we knew Zs, estimation would be trivial

$$\mu_k = \frac{1}{N_k} \sum_{x \in k} x, \quad \Sigma_k = \frac{1}{N_k} \sum_{x \in k} (x - \mu_k)(x - \mu_k)^T$$

What to do?