25/9/19

EE5601: Representation Learning

- Review
- Examples of GMM - multiple populations
- EM for GMM

Recall: $p(\underline{x};\theta) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x; \underline{\mu}_k, \Sigma_k)$; $\sum_{k=1}^{K} \pi_k = 1$; $0 \le \pi_k \le 1$.

Prior: $p(z_k = 1) \equiv \pi_k$; $p(\underline{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$

Posterior: $$p(z_k = 1 \mid \underline{x}) = \frac{p(\underline{x} \mid z_k = 1) \cdot p(z_k = 1)}{p(\underline{x})}$$

$$= \frac{\pi_k \cdot \mathcal{N}(\underline{x}; \underline{\mu}_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_k \mathcal{N}(\underline{x}; \underline{\mu}_j, \Sigma_j)}$$

$$= \gamma(z_k)$$

$\gamma(z_k)$ is the responsibility of the k^{th} mixture comp. in explaining $\underline{x}$

lly $\gamma(z_{nk}) \equiv p(z_k = 1 \mid \underline{x}_n)$.

- Let us find the max likelihood estimates for $\{\pi_k, \underline{\mu}_k, \Sigma_k\}_{k=1}^{K}$
- Assume Σ_k is a diagonal matrix; (for simplified analysis)
- Let us find $\frac{\partial L(X;\theta)}{\partial \mu_{kd}}$ and set it to 0.

Recall $L(X;\theta) = \sum_{n=1}^{N} \log\left[\sum_{j=1}^{K} \pi_j \mathcal{N}(\underline{x}_n; \underline{\mu}_j, \Sigma_j)\right]$.

Also, $\mathcal{N}(\underline{x}; \underline{\mu}_k, \Sigma_k) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_k|}} \exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu}_k)^T \Sigma_k^{-1} \cdot (\underline{x}-\underline{\mu}_k)\right]$

Note that if Σ_k is a diagonal matrix,

$$-\frac{1}{2}(\underline{x}-\underline{\mu}_k)^T \Sigma_k^{-1} \cdot (\underline{x}-\underline{\mu}_k) \Big] = -\frac{1}{2} \sum_{i=1}^{d} \frac{(x_i - \mu_{ki})^2}{\sigma_{ki}^2}$$

Simplifying $\frac{\partial}{\partial \mu_{kd}} L(\underline{x};\theta)$:

$$\frac{\partial}{\partial \mu_{kd}}\left[\sum_{n=1}^{N} \log\left[\sum_{j=1}^{K} \pi_j \mathcal{N}(\underline{x}_n; \underline{\mu}_j, \Sigma_j)\right]\right]$$

$$= \sum_{n=1}^{N} \underbrace{\left[\frac{\pi_k \mathcal{N}(\underline{x}_n; \underline{\mu}_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\underline{x}_n; \mu_j, \Sigma_j)}\right]}_{\gamma(z_{nk})} \cdot \left[2 \cdot \Sigma_k^{-1} \cdot (\underline{x}_n - \underline{\mu}_k)\right] z_k = 0$$