23/9/19 EE5801: Representation Learning

- Review
- GMM
- EM for GMM

Review: • Recall the example of a latent/hidden random variable from the coin example.

- Also recall the GMM distribution:

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

$$\underline{\theta} = \{\pi_k, \underline{\mu}_k, \Sigma_k\}_{k=1}^{K} \quad ; \quad \sum_{k=1}^{K} \pi_k = 1 \; ; \; 0 \leq \pi_k \leq 1$$

Goal: Express $p(\underline{x})$ as a marginal distribution of $p(\underline{x}, \underline{z})$ where $\underline{z}$ is a latent random variable

$$p(\underline{x}) = \sum_{\underline{z}} p(\underline{x}, \underline{z})$$

$$= \sum_{\underline{z}} p(\underline{z}) \cdot p(\underline{x}|\underline{z}) . \qquad - (0)$$

Soln: Defn: $\underline{z} \equiv [0, \ldots 1 \ldots 0]^T$ is a one hot vector of length K $\Rightarrow$ $\underline{z}$ is zero everywhere except at the k^{th} index.

- Let $p(z_k = 1) \equiv \pi_k$.
- observation: $\underline{z}$ can take K different values

- Now, $p(\underline{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$ — ①

- Let us define $p(\underline{x} \mid z_k = 1) = \mathcal{N}(\underline{x}; \underline{\mu}_k, \Sigma_k)$

$$p(\underline{x} \mid \underline{z}) = \prod_{k=1}^{K} \mathcal{N}(\underline{x}; \underline{\mu}_k, \Sigma_k)^{z_k} \quad — ②$$

Plug in ① & ② into ⓪. Note that $\underline{z}$ can take K possible values.

$$\Rightarrow \sum_{\underline{z}} p(\underline{z}) \cdot p(\underline{x} \mid \underline{z}) = \sum_{k=1}^{K} \left[\prod_{i=1}^{K} \pi^{z_i} \prod_{j=1}^{K} \mathcal{N}(\underline{x}; \mu_j, \Sigma_j)^{z_j} \right]$$

$$= \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(\underline{x}; \mu_k, \Sigma_k)$$

Q OK, where is this helpful?

A: ① We can now work with $p(\underline{z})$, $p(\underline{x} \mid \underline{z})$ instead of with $p(\underline{x})$.

② Log likelihood function: $L(\underline{x}; \underline{\theta})$

$$= \sum_{i=1}^{N} \log p(x_i; \underline{\theta})$$

$$= \sum_{i=1}^{N} \log \underbrace{\left[\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i; \mu_k, \Sigma_k) \right]}$$

let us suppose we have 1d data