

Factor Analysis

1 Review of Unsupervised Learning

Recall that in unsupervised learning, we're given m unlabelled training set $x^{(i)} \in \mathbb{R}^n$ ($0 < i < m$) that comes from mixture of Gaussians, and we would like to model the density of $P(x) = \sum_z P(x|z).P(z)$ where $z^{(i)} \in \mathbb{R}^k$ is a latent variable that denote which distribution does $x^{(i)}$ belongs to and thus, we assume there are k distributions.

Here, $z^{(i)} \sim \text{Multinomial}(\phi)$, where $\phi_j \geq 0$ and $\sum_j \phi_j = 1$ and $(X^{(i)}|z^{(i)} = j) \sim \mathcal{N}(\mu_j, \Sigma_j)$

$$\begin{aligned} l(\phi, \mu, \Sigma) &= \sum_{i=1}^m \log p(x^{(i)}, \phi, \mu, \Sigma) \\ &= \sum_{i=1}^m \log \sum_{z^{(i)}=1}^k p(x^{(i)}|z^{(i)}; \mu, \Sigma).p(z^{(i)}; \phi) \end{aligned}$$

The EM algorithm can be applied to fit a mixture model.

$$\begin{aligned} w_j &= P(z^{(i)}|x^{(i)}, \phi, \mu, \Sigma) \\ &= \frac{P(x^{(i)}|z^{(i)} = j).P(z^{(i)} = j)}{\sum_{l=1}^k P(x^{(i)}|z^{(i)} = l).P(z^{(i)} = l)} \\ 1. \text{ E- step: } &= \frac{\frac{1}{\sigma\sqrt{2\pi}} \exp[(x^{(i)} - \mu_j)^T \cdot \sum_j^{-1} \cdot (x^{(i)} - \mu_j)] \cdot \phi_j}{\sum_{l=1}^k (\frac{1}{\sigma\sqrt{2\pi}} \exp[(x^{(i)} - \mu_j^{(i)})^T \cdot \sum_j^{-1} \cdot (x^{(i)} - \mu_j^{(i)})])} \\ 2. \text{ M- step: } &\phi_j = \frac{1}{m} \sum_{i=1}^m w_j^{(i)}; \Sigma_j = \frac{\sum_{i=1}^m w_j^{(i)} \cdot (x^{(i)} - \mu_j) \cdot (x^{(i)} - \mu_j)^T}{\sum_{i=1}^m w_j^{(i)}} \\ \mu_j &= \frac{\sum_{i=1}^m w_j^{(i)} \cdot x^{(i)}}{\sum_{i=1}^m w_j^{(i)}}; \end{aligned}$$

1.1 Problem Statement

Now suppose that we have $n \gg m$, we will find that Σ is singular matrix. This means Σ^{-1} doesn't exist and we find $1/|\Sigma|^{\frac{1}{2}} = 1/0$. Those terms are needed to compute EM algorithm (refer to E-step).