Minimum error formulation of PCA

$x_1, \ldots, x_N \in \mathbb{R}^D$

Want to find basis vectors $\{u_i\}$. $\left\{ u_i^T u_j = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases} \right.$

$$\underline{\underline{\tilde{x}_n}} = \sum_{i=1}^{M} \underline{\underline{z_{ni}}}\, u_i + \sum_{i=M+1}^{D} \underline{\underline{b_i}}\, u_i$$

We want to minimize

$$J = \frac{1}{N} \sum_{n=1}^{N} \| x_n - \tilde{x}_n \|_2^2$$

Find the optimal z_{ni}, b_i and u_i

① Let's find optimal z_{nj}

$u_i^T u_j \quad i \neq j$

$$J = \frac{1}{N} \cancel{\sum_{n=1}^{N}} \| x_n - \sum_{i=1}^{M} z_{ni} u_i - \sum_{i=M+1}^{D} b_i u_i \|^2$$

$$\|a+b+c\|^2 = \|a\|^2 + \|b\|^2 + \|c\|^2 + 2a^Tb + 2a^Tc + 2b^Tc$$

$$J = \text{const.} + z_{nj}^2 \cancel{u_j^T u_j} - 2 z_{nj} u_j^T x_n$$

$$\frac{\partial J}{\partial z_{nj}} = 0 = 2z_{nj} - 2\cancel{z_{nj}}\, u_j^T x_n \Rightarrow z_{nj} = x_n^T \underline{\underline{u_j}}$$

② Find optimal b_j

$$J = \frac{1}{N} \sum_{n=1}^{N} \| x_n - \sum_{i=1}^{M} z_{ni} u_i - \sum_{i=M+1}^{D} b_i u_i \|^2$$

$$J = \text{const.} + \frac{1}{N} \sum_{n=1}^{N} \left[b_j^2 - b_j u_j^T x_n \right]$$

$$\frac{\partial J}{\partial b_j} = 0 = 2b_j - 2\frac{1}{N} \sum_{n=1}^{N} x_n^T u_j \Rightarrow \underline{\underline{b_j = \frac{1}{N} \sum_{n=1}^{N} x_n^T u_j}} = \bar{x}^T u_j$$

③ Find $\{u_i\}$

$$\tilde{x}_n = \sum_{i=1}^{M} (x_n^T u_i) u_i + \sum_{i=M+1}^{D} (\underline{\bar{x}^T u_i}) u_i$$

$$x_n = \sum_{i=1}^{D} (\underline{x_n^T u_i}) \cdot u_i$$

$$J = \frac{1}{N} \sum_{n=1}^{N} \| x_n - \tilde{x}_n \|^2$$

$$= \frac{1}{N} \sum_{n=1}^{N} \Big\| \sum_{i=M+1}^{D} \underbrace{(x_n - \bar{x})^T u_i} \; u_i \Big\|_2^2 \qquad S = \frac{1}{N} X^T X$$

$$= \frac{1}{N} \sum_{n=1}^{N} \sum_{i=M+1}^{D} \big(\underbrace{x_n^T u_i}_{\text{projeted data}} - \underbrace{\bar{x}^T u_i}_{\text{mean projeted data.}} \big)^2 = \sum_{i=M+1}^{D} u_i^T S u_i$$

In this minimum error formulation.

We want $\min \sum_{i=M+1}^{D} u_i^T S u_i$

In the maximum error formulation.

We want $\max \sum_{i=1}^{M} u_i^T S u_i$

2. E-M.

log ~~of~~ joint probability

$$L = \sum_{n=1}^{N} \log\left(\sum_{k=1}^{K} \pi_k N(x_n; \mu_k, \Sigma_k) \right)$$

$$\underbrace{\frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_n-\mu_k)^T \Sigma_k^{-1} (x_n-\mu_k)}}$$

Want to find optimal for μ_l

$$\frac{\partial L}{\partial \mu_l} = \sum_{n=1}^{N} \frac{\pi_l}{\sum_{k=1}^{K} \pi_k N(x_n; \mu_k, \Sigma_k)} \frac{\partial}{\partial \mu_l}$$

$N(x_n; \mu_l, \Sigma_l)$ //

$$\frac{\partial L}{\partial \mu_l} = \sum_{n=1}^{N} \frac{\pi_l}{\sum_{k=1}^{K} \pi_k N(x_n; \mu_k, \Sigma_k)} \left(\frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_l|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_n-\mu_l)^T \Sigma_k^{-1} (x_n-\mu_l)} \right) \frac{\partial}{\partial \mu_l}\left[-\frac{1}{2}(x_n-\mu_l)^T \Sigma_l^{-1} (x_n-\mu_l) \right]$$

γ_{nl}

$$\frac{\partial L}{\partial \mu_l} = \sum_{n=1}^{N} \frac{\pi_l N(x_n; \mu_l, \Sigma_l)}{\sum_{k=1}^{K} \pi_k N(x_n; \mu_k, \Sigma_k)} \left(-\frac{1}{2}\right) \cdot (-1) \cdot 2 \Sigma_l^{-1} (x_n - \underline{\mu_l}) = 0$$

3. Special problem for the Gaussian mixture model.

Singularity issue.

Suppose we model data as a mixture of 2-univariate Gaussian,

$\cancel{\pi_1}$ $\cancel{\pi_2}$ μ_1 $\cancel{\mu_2}$ σ_1 $\cancel{\sigma_2}$

maximize

$$P = \prod_{i=1}^{N} \left[\underbrace{\cancel{\pi_1} \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{1}{\sigma_1^2}(x_n-\mu_1)^2}}_{\to 0} + \underbrace{\cancel{\pi_2 \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{1}{\sigma_2^2}(x_n-\mu_2)^2}}}_{>0} \right]$$

Consider the case that at one step, we have $\mu_1 = x_j$

$$P = \underbrace{\prod_{\substack{i=1 \\ i\neq j}}^{N} \left[\quad \right]}_{\to 0} \times \underbrace{\left[\pi \underbrace{\frac{1}{\sqrt{2\pi}\,\sigma_1}}_{\to \infty} e + \quad \right]}_{\to 0} \xrightarrow{\sigma_1 \to 0} \infty$$

$\sigma_1 \to 0$

Use hurestic to avoid singularity

We can detect when $\mu_1, \mu_2 = x_n$

and then reset μ_1 μ_2