Overview

- Brush Factor Analysis
- PCA

Factor Analysis Recap.

$x^{(1)}, x^{(2)} \cdots x^{(n)} \in \mathbb{R}^d$, $n \ll d$

Recall likelihood function

$$l(\mu, \Sigma) = \sum_{i=1}^{n} \log \frac{1}{2\pi |\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2} (x^{(i)} - \mu)^T \Sigma^{-1} (x - \mu) \right\}$$

$$= \sum_{i=1}^{n} -\frac{1}{2} (x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu) - \log 2\pi |\Sigma|^{1/2}$$

Hence its equivilant to minimize

$$g(\mu, \Sigma) = \sum_{i=1}^{n} (x - \mu)^T \Sigma^{-1} (x - \mu) + \log |\Sigma|$$

we recall restrictions on Σ

- Rank $n \ll d$, Σ^{-1} is not defined, $|\Sigma| = 0$

Estimate of Σ is full rank, easy to estimate μ

$$\nabla_\mu g(\mu, \Sigma) = 2 \sum_{i=1}^{n} \Sigma^{-1} (x^{(i)} - \mu) = 0 \quad \text{(critical points)}$$

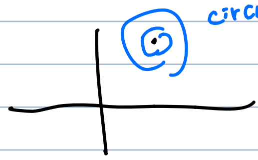
$$\Leftrightarrow \sum_{i=1}^{n} x^{(i)} - \mu = 0 \quad (\Sigma \text{ is full rank})$$

$$\Rightarrow \mu = \frac{1}{n} \sum_{i=1}^{n} x^{(i)}$$

$\therefore$ μ is easy to estimate if Σ is full rank.

Building block 1: $\Sigma = \sigma^2 I$, σ^2 is just a scalar

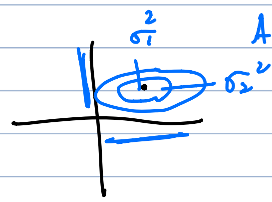
circular covariance



last time: $\sigma^2 = \frac{1}{nd} \| x^{(i)} - \mu \|^2$

Building block 2:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & & \ddots \ \sigma_n^2 \end{bmatrix}$$



Axis aligned.

$$g(\mu, \Sigma) = \sum_{i=1}^{n} (x-\mu)^T \Sigma^{-1} (x-\mu) + \log|\Sigma| \qquad |\Sigma| = \prod_{j=1}^{d} \sigma_j^2$$

$z_i \triangleq \sigma_i^2$

$$= \sum_{i=1}^{n} \sum_{j=1}^{d} z_j^{-2} (x_j^{(i)} - \mu_j)^2 + \log z_j$$

Decode into d problems.

fix $j \Rightarrow \sum_{i=1}^{n} z_j^{-2} (x_j^{(i)} - \mu_j)^2 + \log z_j$ (derivative, set to zero)

$$\Rightarrow z_j = \sigma_j^2 = \frac{1}{n} \sum_{i=1}^{n} (x^{(i)} - \mu_j)^2$$

$\therefore$ We assume that this axis-aligned structure could fit the component

Our Factor Model

Parameters

fit $\begin{cases} \mu \in \mathbb{R}^d \\ \Lambda \in \mathbb{R}^{d\times s} \\ \Phi \in \mathbb{R}^{d\times d} \end{cases}$ $s < d$, "small dim"

Model

$P(x, z) = P(x|z)P(z)$ z is latent variable

$z \sim N(0, I) \in \mathbb{R}^s$

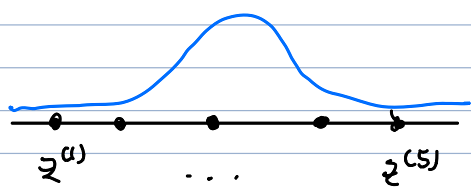
$x = \mu + \Lambda z + \varepsilon$ on $x \sim N(\mu + \Lambda s, \Phi)$

mean ↲ ↳ map from small to high dim

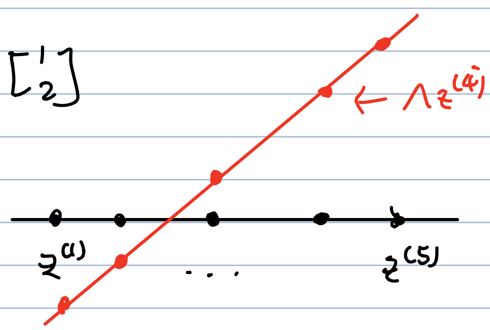
$\varepsilon \sim N(0, \Phi)$

Ex: $d = 2$, $s = 1$, $n = 5$: $x = \mu + \Lambda z + \varepsilon$

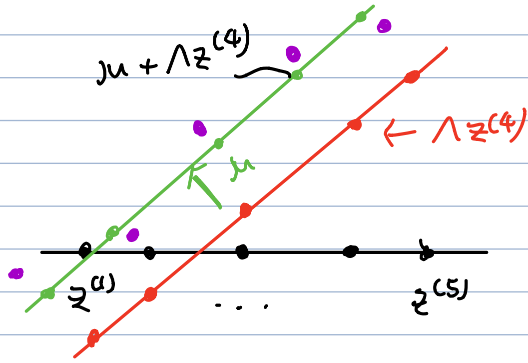
① Generate $z^{(1)}, z^{(2)}, \ldots z^{(5)}$ from $N(0, 1)$



② $\Lambda = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ Map to line



③



$x = \mu + \Lambda z + \epsilon$

Blows up the dim.

Technical Tools

Block Gaussians

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad x_1 \in \mathbb{R}^{d_1} \qquad x_2 \in \mathbb{R}^{d_2} \qquad x \in \mathbb{R}^{d_1 + d_2}$$

$$\Sigma = \begin{array}{cc} d_1 & d_2 \end{array}$$
$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{array}{l} \} d_1 \\ \} d_2 \end{array} \qquad \Sigma_{ij} \in \mathbb{R}^{d_i \times d_j} \qquad ij \in \{1, 2\}$$

Fact 1. $P(x_1) = \int_{x_2} P(x_1, x_2)$ Marginalization

For gaussians $p(x_1) = \mathcal{N}(\mu_{1}, \Sigma_{11})$ Gaussian

Fact 2 $P(x_1 | x_2) \sim \mathcal{N}(\mu_{1|2}, \Sigma_{1|2})$ conditions

$$\mu_{1|2} = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2)$$
$$\Sigma_{1|2} = \Sigma_{11}$$

Matrix Inversion

Factor Analysis

$$x = \mu + \Lambda z + \varepsilon$$

$$\begin{pmatrix} z \\ x \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ \mu \end{pmatrix}, \Sigma \right)$$

Since $\mathbb{E}[z] = 0$
$\mathbb{E}[x] = \mu$

$$\Sigma_{11} = \mathbb{E}[z \cdot z^T] = I \qquad (z \sim \mathcal{N}(0, I))$$

$$\Sigma_{12} = \mathbb{E}[z(x - \mu)^T] = \mathbb{E}[\underbrace{z z^T}_{I} \Lambda^T] + \mathbb{E}[\underbrace{\cancel{z}}_{0} \epsilon^T] = \Lambda^T$$

$$\Sigma_{21} = \Sigma_{12}^T \quad (\Sigma \text{ is symmetric})$$

$$\Sigma_{22} = \mathbb{E}[(x - \mu)(x - \mu)^T] = \mathbb{E}[(\Lambda z + \varepsilon)(\Lambda z + \varepsilon)^T]$$
$$= \mathbb{E}[\Lambda z z^T \Lambda^T] + \mathbb{E}[\epsilon \epsilon^T] = \Lambda \Lambda^T + \Phi$$

$$\Sigma = \begin{bmatrix} I & \Lambda^T \\ \Lambda & \Lambda \Lambda^T + \Phi \end{bmatrix}$$

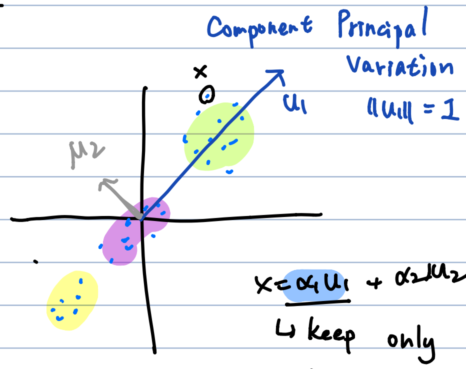
EM Algorithm

E Step : $Q_i(z^{(i)}) = P(z^{(i)} \mid x^{(i)}, \theta)$ - use conditional

M Step : We know how to estimate

PCA : Principle Component Analysis

Structure	Prob	Non Prob
clusters	GMM Factor	K-means
Subspace	Factor Analysis	PCA



$x^{(1)}, x^{(2)}, \ldots, x^{(n)} \in \mathbb{R}^d$

$\mu = \frac{1}{n} \sum x^{(i)}$

$x^{(i)} \mapsto x^{(i)} - \mu$ (centering)

$x = \alpha_1 u_1 + \alpha_2 u_2$

↳ keep only 1 term to compress data