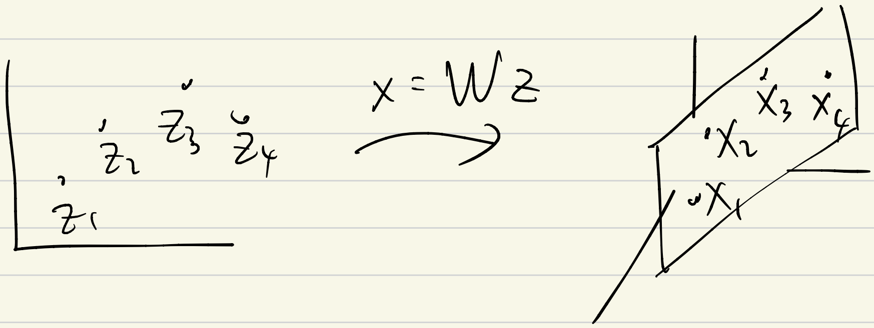
Probabilistic PCA

Yung-Kyun Noh

Principal Component Analysis

$z \in \mathbb{R}^d, \quad x \in \mathbb{R}^D \qquad D > d$



$W \in \mathbb{R}^{D \times d}$

Projection matrix

$\mathcal{D} = \{x_i\}_{i=1}^{N} \quad x_i \in \mathbb{R}^D$

$$L = \sum_{i=1}^{N} \| x_i - W z_i \|^2 = \mathrm{tr}\left[(X - WZ)^T (X - WZ) \right]$$

$$Z = \begin{pmatrix} | & & | \\ z_1 & \cdots & z_N \\ | & & | \end{pmatrix}, \quad X = \begin{pmatrix} | & & | \\ x_1 & \cdots & x_N \\ | & & | \end{pmatrix}$$

$$\left(\frac{d}{dA} \mathrm{tr}[A^T B] = B, \qquad \mathrm{tr}[ABC] = \mathrm{tr}[BCA] \right)$$

$$\frac{dL}{dZ} = W^T (X - WZ) = 0$$

$$W^T X - W^T W Z = 0$$

$$Z^* = (W^T W)^{-1} W^T X$$

$$L = \frac{1}{2} tr[(X - WZ)^T (X - WZ)]$$

$$= \frac{1}{2} tr[(X - W(W^TW)^{-1}W^TX)^T (X - W(W^TW)^{-1}W^TX)]$$

$$= \frac{1}{2} tr[X^T(I - W^T(W^TW)^{-1}W)^T (I - W(W^TW)^{-1}W) X]$$

$$= \frac{1}{2} tr[X^T(I - 2W(W^TW)^{-1}W + W(W^TW)^{-1}W) X]$$

$$= \frac{1}{2} tr[X^T(I - W(W^TW)^{-1}W^T) X]$$

$$= \frac{1}{2} tr[XX^T] - \frac{1}{2} tr[(W^TW)^{-1}W^TXX^TW]$$

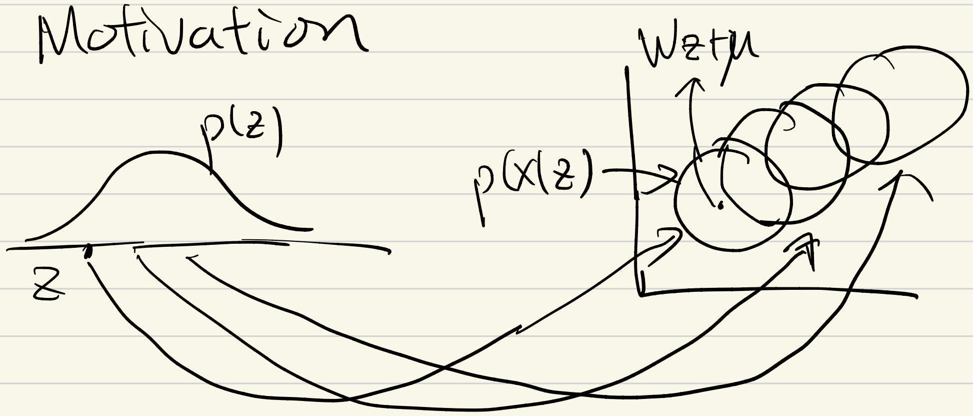
$$W^* = \underset{W}{\operatorname{argmax}} \; tr[(W^TW)^{-1}W^TXX^TW]$$

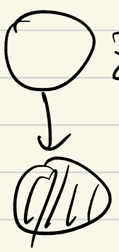
$$\Rightarrow W^* = \begin{pmatrix} | & & | \\ w_1 & \cdots & w_d \\ | & & | \end{pmatrix}$$ where $w_1, \cdots, w_d$ are the eigenvectors having maximum eigenvalues of XX^T.

If we consider centralized data $X = \begin{pmatrix} | & & | \\ x_1 - \mu, & \cdots & x_N - \mu \\ | & & | \end{pmatrix}$ with $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$, then $C = \frac{1}{N}XX^T$ is the covariance matrix. Therefore $w_1, \cdots, w_d \in \mathbb{R}^D$ are the eigenvectors of the covariance matrix with corresponding largest eigenvalues.

Probabilistic PCA

Motivation





$z \in \mathbb{R}^d$ $\qquad p(z) = \mathcal{N}(z \mid 0, I)$

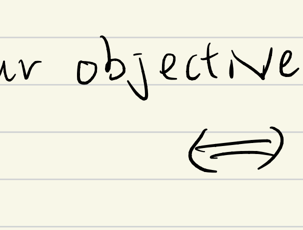
$x \in \mathbb{R}^D$ $\qquad p(x|z) = \mathcal{N}(x \mid Wz+\mu, \sigma^2 I)$

noise

$W \in \mathbb{R}^{D \times d}$

$\mu \in \mathbb{R}^D$

Our information is about $p(x)$ $\Leftarrow$ Marginalize z.



$p(x) = \int p(x|z)p(z)\,dz$

x: observed data

Our objective: Find parsimoneous representation

$\Longleftrightarrow$ Find $p(z|x)$

z: latent low-dimensional parsimoneous representation

x: data

Plan: $p(z)$, $p(x|z) \Rightarrow$ Joint $p(x,z) \xRightarrow{\text{marginalize}}$ Get $p(x)$

Model (Both are Gaussians)

$p(z|x)$ with estimated parameters $\Leftarrow$ Estimate parameters

Find a map z_i for x_i for $i=1,\cdots,N$

Model: $\begin{cases} p(z) = \dfrac{1}{\sqrt{2\pi}^d} \exp\left(-\dfrac{1}{2}z^2\right) \\ p(x|z) = \dfrac{1}{\sqrt{2\pi\sigma^2}^D} \exp\left(-\dfrac{1}{2\sigma^2}(x-\underline{Wz-\mu})^2\right) \end{cases}$

$Wz+\mu$: center of sphere

Joint $p(x,z)$

$= p(z)p(x|z)$

$= \dfrac{1}{\sqrt{2\pi}^{D+d}\sigma^D} \exp\left(-\dfrac{1}{2}z^2 - \dfrac{1}{2\sigma^2}(x-Wz-\mu)^2\right)$

Find quadratic form of $\begin{pmatrix} z \\ x \end{pmatrix}$ inside exp

$$z^2 + \frac{x^2}{\sigma^2} - \frac{2}{\sigma^2}x^T(Wz+\mu) + \frac{1}{\sigma^2}(Wz+\mu)^2$$

$$= \left[\begin{pmatrix} z \\ x \end{pmatrix} - \mu_{x,z}\right]^T \Sigma_{x,z}^{-1} \left[\begin{pmatrix} z \\ x \end{pmatrix} - \mu_{x,z}\right]$$

$$z^2 + \frac{z^T W^T W z}{\sigma^2} + \frac{x^2}{\sigma^2} - \frac{2x^T W z}{\sigma^2} - \frac{2x^T \mu}{\sigma^2} + \frac{2\mu^T W z}{\sigma^2} + \cdots$$

$$= \begin{pmatrix} z \\ x \end{pmatrix}^T \begin{bmatrix} I + \frac{W^T W}{\sigma^2} & -\frac{W^T}{\sigma^2} \\ -\frac{W}{\sigma^2} & \frac{I}{\sigma^2} \end{bmatrix} \begin{pmatrix} z \\ x \end{pmatrix} - \frac{2}{\sigma^2} \begin{pmatrix} z \\ x \end{pmatrix}^T \begin{pmatrix} -W^T \mu \\ \mu \end{pmatrix} + \cdots$$

$$\left(\begin{aligned} \Sigma_{x,z}^{-1} &= \begin{pmatrix} I + \frac{W^T W}{\sigma^2} & -\frac{W^T}{\sigma^2} \\ -\frac{W}{\sigma^2} & \frac{I}{\sigma^2} \end{pmatrix} = \frac{1}{\sigma^2} \begin{pmatrix} W^T W + \sigma^2 I & -W^T \\ -W & I \end{pmatrix} \\ \mu_{x,z} &= \Sigma_{x,z} \cdot \frac{1}{\sigma^2} \begin{pmatrix} W^T \mu \\ -\mu \end{pmatrix} = \begin{pmatrix} W^T W + \sigma^2 I & -W^T \\ -W & I \end{pmatrix}^{-1} \begin{pmatrix} W^T \mu \\ -\mu \end{pmatrix} \end{aligned} \right.$$

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1} + D^{-1}CMBD^{-1} \end{pmatrix} \searrow \Sigma_{x,z}$$

$$M = (A - BD^{-1}C)^{-1}$$

M in $\Sigma_{x,z}$ is
$$\begin{aligned} M &= \{A - BD^{-1}C\}^{-1} \\ &= \{W^T W + \sigma^2 I - (-W^T) I (-W)\}^{-1} \\ &= (\sigma^2 I)^{-1} = \frac{1}{\sigma^2} I \end{aligned}$$

$$\Rightarrow \begin{cases} \Sigma_{x,z} = \sigma^2 \begin{pmatrix} \frac{1}{\sigma^2} I & -\frac{1}{\sigma^2} I (-W^T) I \\ -I(-W)\frac{1}{\sigma^2} I & I + I(-W)\frac{1}{\sigma^2} I (-W^T) I \end{pmatrix} = \begin{pmatrix} I & W^T \\ W & WW^T + \sigma^2 I \end{pmatrix} \\ \mu_{x,z} = \frac{1}{\sigma^2} \Sigma_{x,z} \begin{pmatrix} -W^T \mu \\ \mu \end{pmatrix} = \begin{pmatrix} 0 \\ \mu \end{pmatrix} \end{cases} \Rightarrow \text{Joint density}$$

$$\Sigma_{x,z} = \begin{pmatrix} I & W^T \\ W & \underbrace{WW^T + \sigma^2 I}_{\Sigma_x} \end{pmatrix} \qquad \mu_{x,z} = \begin{pmatrix} 0 \\ \underbrace{\mu}_{\mu_x} \end{pmatrix} \begin{matrix} \leftarrow z \\ \leftarrow x \end{matrix}$$

Estimate from data

$$W_{ML} = U_d (L_d - \sigma^2 I_d)^{\frac{1}{2}} R$$

$$U_d = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_d \\ | & & | \end{pmatrix}, \quad L_d = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{pmatrix}$$ with the eigenvectors $u_1, \cdots, u_d$ corresponding to the maximum d eigenvalues $\lambda_1, \cdots, \lambda_d$, and $R \in \mathbb{R}^{d \times d}$ is an orthonormal matrix.

$$\left(\text{Note } \left. \frac{dp(x; W)}{dW} \right|_{W_{ML}} = 0 \right)$$

If necessary, σ^2_{ML} can be obtained similarly

$$\sigma^2_{ML} = \frac{1}{D-d} \sum_{i=d+1}^{D} \lambda_i$$

Conditioning

$$\Rightarrow \quad p(z|x) = N\left(z \,\middle|\, \mu_z + \Sigma_{zx} \Sigma_x^{-1} (x - \mu_x), \; \Sigma_z - \Sigma_{zx} \Sigma_x^{-1} \Sigma_{xz} \right)$$

$$= N\left(z \,\middle|\, 0 + W_{ML}^T (W_{ML} W_{ML}^T + \sigma^2 I)^{-1} (x - \mu), \; I - W_{ML}^T (W_{ML} W_{ML}^T + \sigma^2 I)^{-1} W_{ML} \right)$$

$$\therefore \quad z_i = \left(W_{ML}^T W_{ML} + \sigma^2 I_d\right)^{-1} W_{ML}^T (x_i - \mu)$$