

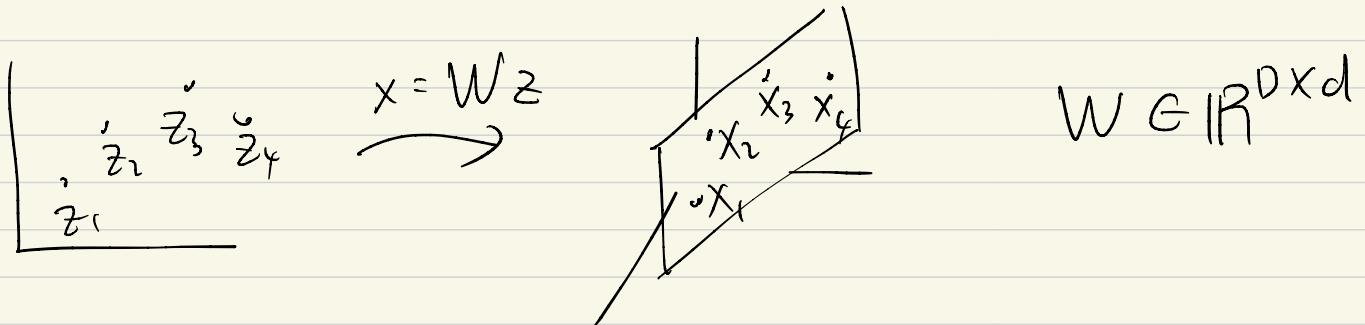
Probabilistic PCA & GPLVM

Yung-Kyun Noh



Principal Component Analysis

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



Projection matrix

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

$$L = \sum_{i=1}^N \|x_i - Wz_i\|^2 = \text{tr}[(X - WZ)^T(X - WZ)]$$

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

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

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

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

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

$$\begin{aligned}
L &= \frac{1}{2} \text{tr} \left[(X - WZ)^T (X - WZ) \right] \\
&= \frac{1}{2} \text{tr} \left[(X - W(W^T W)^{-1} W^T X)^T (X - W(W^T W)^{-1} W^T X) \right] \\
&= \frac{1}{2} \text{tr} \left[X^T (I - W(W^T W)^{-1} W)^T (I - W(W^T W)^{-1} W) X \right] \\
&= \frac{1}{2} \text{tr} \left[X^T (I - 2W(W^T W)^{-1} W + W(W^T W)^{-1} W) X \right] \\
&= \frac{1}{2} \text{tr} \left[X^T (I - W(W^T W)^{-1} W^T) X \right] \\
&= \frac{1}{2} \text{tr} [XX^T] - \frac{1}{2} \text{tr} [(W^T W)^{-1} W^T XX^T W]
\end{aligned}$$

$$W^* = \underset{W}{\operatorname{argmax}} \text{tr} \left[(W^T W)^{-1} W^T XX^T W \right]$$

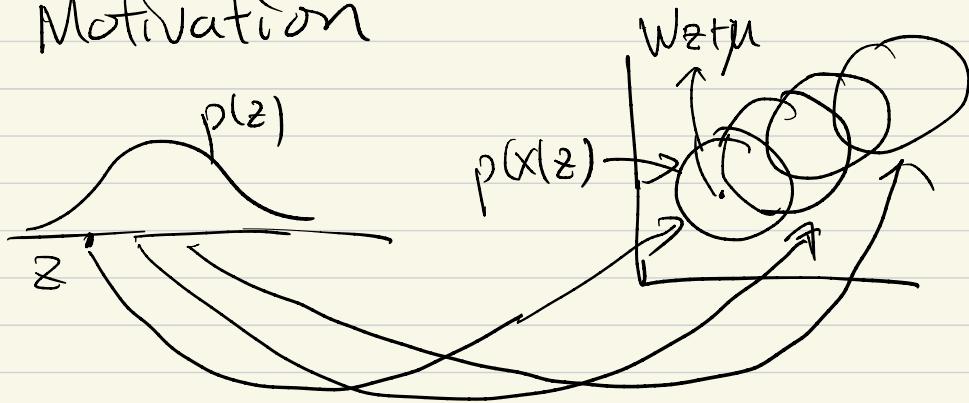
$$\Rightarrow W^* = \begin{pmatrix} w_1 & \cdots & w_d \\ \vdots & \ddots & \vdots \end{pmatrix} \text{ where } w_1, \dots, w_d \text{ are}$$

the eigenvectors having maximum eigenvalues of XX^T .

If we consider centralized data $X = (x_1 - \mu, \dots, x_n - \mu)$
with $\mu = \frac{1}{n} \sum_{i=1}^n x_i$, then $C = \frac{1}{n} XX^T$ is the covariance matrix. Therefore $w_1, \dots, w_d \in \mathbb{R}^D$ are the eigenvectors of the covariance matrix with corresponding largest eigenvalues.

Probabilistic PCA.

Motivation



$$z \in \mathbb{R}^d$$

$$p(z) = N(z | 0, I)$$

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

$$p(x|z) = N(x | Wz + \mu, \sigma^2 I)$$

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

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

↑
noise

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

$$\boxed{\int_{\mathbb{R}^d} p(x) = \int p(x|z)p(z)dz}$$

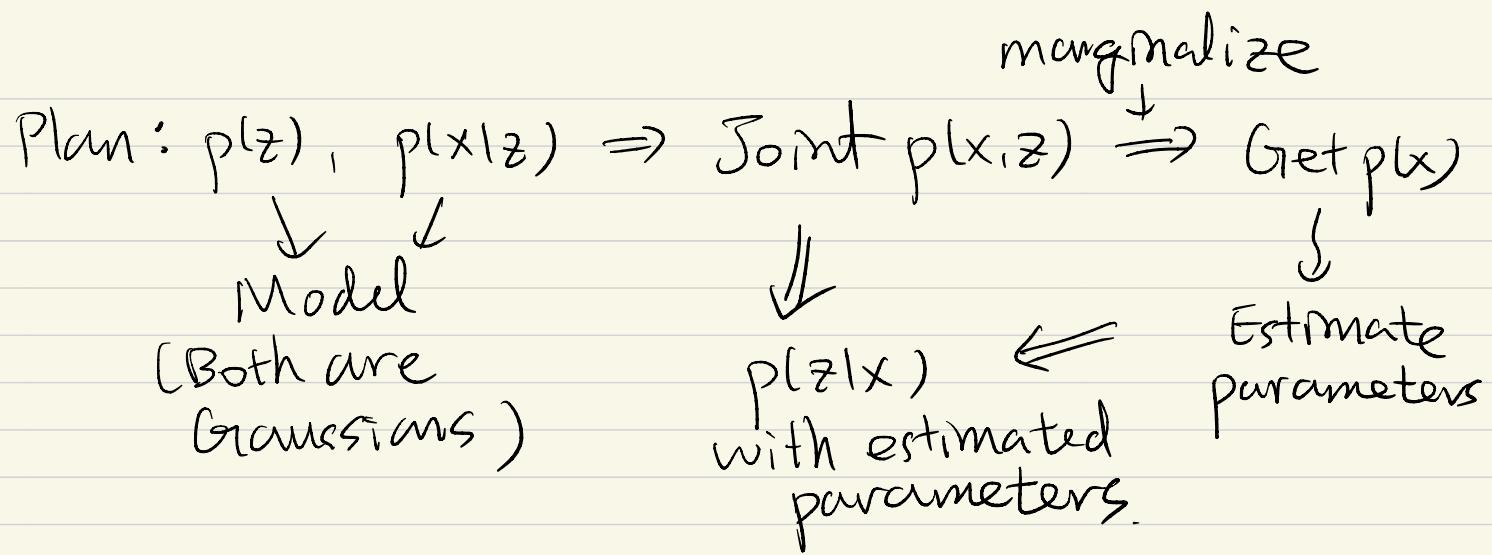
x: observed data

Our objective: Find parsimonious representation

$$\Leftrightarrow \text{Find } p(z|x)$$

\downarrow data

latent
low-dimensional
parsimonious representation



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

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

Joint $p(x,z)$

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

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

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

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

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

$Wz + \mu$:
center of sphere

$$z^2 + \frac{z^T W^T W z}{6^2} + \frac{x^2}{6^2} - \frac{2x^T W z}{6^2} - \frac{2x^T \mu}{6^2} + \frac{2\mu^T W z}{6^2} + \dots$$

$$= \begin{pmatrix} z \\ x \end{pmatrix}^T \begin{bmatrix} I + \frac{W^T W}{6^2} & -\frac{W^T}{6^2} \\ -\frac{W}{6^2} & \frac{I}{6^2} \end{bmatrix} \begin{pmatrix} z \\ x \end{pmatrix} - \frac{2}{6^2} \begin{pmatrix} z \\ x \end{pmatrix}^T \begin{pmatrix} -W^T \mu \\ \mu \end{pmatrix} + \dots$$

$$\left(\begin{array}{l} \Sigma_{x,z}^{-1} = \begin{pmatrix} I + \frac{W^T W}{6^2} & -\frac{W^T}{6^2} \\ -\frac{W}{6^2} & \frac{I}{6^2} \end{pmatrix} = \frac{1}{6^2} \begin{pmatrix} W^T W + 6^2 I & -W^T \\ -W & I \end{pmatrix} \\ M_{x,z} = \Sigma_{x,z} \cdot \frac{1}{6^2} \begin{pmatrix} W^T \mu \\ -\mu \end{pmatrix} = \begin{pmatrix} W^T W + 6^2 I & -W^T \\ -W & I \end{pmatrix}^{-1} \begin{pmatrix} W^T \mu \\ -\mu \end{pmatrix} \end{array} \right)$$

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

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

$$M \text{ in } \Sigma_{x,z} \text{ is } M = \{A - BD^{-1}C\}^{-1}$$

$$= \{W^T W + 6^2 I - (-W^T) I (-W)\}^{-1}$$

$$= (6^2 I)^{-1} = \frac{1}{6^2} I$$

$$\Rightarrow \left\{ \begin{array}{l} \Sigma_{x,z} = 6^2 \begin{pmatrix} \frac{1}{6^2} I & -\frac{1}{6^2} I (-W^T) I \\ -I (-W) \frac{1}{6^2} I & I + I (-W) \frac{1}{6^2} I (-W^T) I \end{pmatrix} = \begin{pmatrix} I & W^T \\ W & WW^T + 6^2 I \end{pmatrix} \\ M_{x,z} = \frac{1}{6^2} \Sigma_{x,z} \begin{pmatrix} -W^T \mu \\ \mu \end{pmatrix} = \begin{pmatrix} 0 \\ \mu \end{pmatrix} \Rightarrow \text{Joint density} \end{array} \right.$$

$$\Sigma_{x,z} = \begin{pmatrix} I & W^T \\ W & \underbrace{WW^T + \sigma^2 I}_{\Sigma_x} \end{pmatrix}$$

$$\mu_{x,z} = \begin{pmatrix} 0 \\ \mu \\ \vdots \\ \mu \end{pmatrix} \leftarrow \begin{matrix} z \\ x \end{matrix}$$

Estimate from data

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

$$U_d = \begin{pmatrix} u_1 & \cdots & u_d \end{pmatrix}, L_d = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{pmatrix} \text{ with the}$$

eigenvectors u_1, \dots, u_d corresponding to the maximum d eigenvalues $\lambda_1, \dots, \lambda_d$, and $R \in \mathbb{R}^{d \times d}$ is an orthonormal matrix,

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

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

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

Conditioning

$$\begin{aligned} \Rightarrow p(z|x) &= N(z | \mu_z + \Sigma_{zx} \Sigma_x^{-1} (x - \mu_x), \\ &\quad \Sigma_z - \Sigma_{zx} \Sigma_x^{-1} \Sigma_{xz}) \\ &= N(z | 0 + W_{ML}^T (WW_{ML}^T + \sigma^2 I)^{-1} (x - \mu), \\ &\quad I - W_{ML}^T (WW_{ML}^T + \sigma^2 I)^{-1} W_{ML}) \end{aligned}$$

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

GPLVM

Z is given, W has density function.

The randomness of data X comes from the randomness of W .

$$P(X|Z, W) = \prod_n p(x_n|z_n, W)$$

$$= \frac{1}{\sqrt{2\pi}^D \beta^{-\frac{DN}{2}}} \exp\left(-\frac{\beta}{2} \sum_n (x_n - Wz_n)^2\right)$$

$$\sum_i (x_{ni} - w_i^T z_n)^2$$

$$W = \begin{pmatrix} -w_1^T - \\ \vdots \\ -w_D^T - \end{pmatrix} \in \mathbb{R}^{D \times d}$$

$$x_n = Wz_n$$

$$x_{ni} = w_i^T z_n$$

(Features are independent from each other
& Data are independent from each other,

$$P(W) = \frac{1}{\sqrt{2\pi}^d \alpha^{-\frac{dD}{2}}} \exp\left(-\frac{\alpha}{2} \sum_{i=1}^D w_i^2\right)$$

All scalar elements
are independent.

Marginalize W

$$p(X|Z, W) p(W) = \frac{1}{\sqrt{2\pi}^{DN+dD} \alpha^{-\frac{dD}{2}} \beta^{-\frac{DN}{2}}} \exp\left(-\frac{\beta}{2} \sum_n (x_n - Wz_n)^2 - \frac{\alpha}{2} \sum_i w_i^2\right)$$

$$\alpha \sum_i w_i^2 + \beta \sum_n (x_n - Wz_n)^2 \rightarrow \text{square of each element.}$$

$$= \sum_{i=1}^D \left\{ \alpha w_i^T w_i + \sum_n \beta (z_n^T w_i w_i^T z_n - 2x_{ni} w_i^T z_n + x_{ni}^2) \right\}$$

$$w_i^T \sum_n z_n z_n^T w_i$$

$$\begin{aligned}
 & \stackrel{d \times d}{\downarrow} \quad \downarrow z = (z_1, \dots, z_n) \\
 = \sum_{i=1}^D & \left\{ w_i^T (\alpha I + \beta Z Z^T) w_i \right. \\
 & - 2\beta w_i^T (\alpha I + \beta Z Z^T) (\alpha I + \beta Z Z^T)^{-1} Z X_i \\
 & + \beta \sum_n x_{ni}^2 \\
 & \left. + \beta \sum_n x_{ni}^2 - \beta^2 X_{::i}^T Z^T (\alpha I + \beta Z Z^T)^{-1} Z X_{::i} \right\}
 \end{aligned}$$



$$\sum_{i=1}^D \beta X_{::i}^T (I_N - \beta Z^T (\alpha I + \beta Z Z^T)^{-1} Z) X_{::i}$$

$$= \sum_{i=1}^D \beta X_{::i}^T (I_N + \frac{\beta}{\alpha} Z^T Z)^{-1} X_{::i}$$

$$= \sum_{i=1}^D X_{::i}^T (\beta^{-1} I_N + \alpha^{-1} Z^T Z)^{-1} X_{::i}$$

$$= \text{tr} [(\beta^{-1} I_N + \alpha^{-1} Z^T Z)^{-1} X^T X]$$

$$X = \begin{pmatrix} | & | \\ X_1 & \dots & X_D \\ | & | \end{pmatrix} = \begin{pmatrix} -X_{::1}^T & - \\ \vdots & \vdots \\ -X_{::D}^T & - \end{pmatrix}$$

$$X_{::i} \in \mathbb{R}^N$$

$$\begin{aligned}
 & (A + B D^{-1} C)^{-1} \\
 & = A^{-1} - A^{-1} B (D + C A^{-1} B)^{-1} C A^{-1}
 \end{aligned}$$

$$A = I, \quad B = \sqrt{\beta} Z^T$$

$$C = \sqrt{\beta} Z, \quad D = \alpha I$$

GP extension:

$$\text{maximize } \text{tr} [(\beta^{-1} I_N + \alpha^{-1} Z^T Z)^{-1} K]$$

with respect to Z . Use gradient ascent!