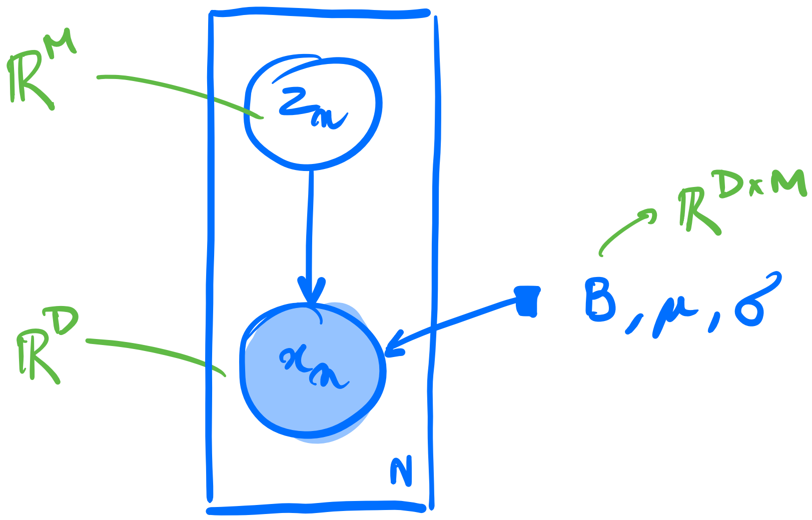
LEC 03

Probabilistic PCA



$$z_n \sim \mathcal{N}(z \mid 0, I)$$

$$x_n \mid z_n \sim \mathcal{N}(x \mid B z_n + \mu, \sigma^2 I)$$

$$p(z) = \mathcal{N}(z \mid 0, I)$$

$$p(x \mid z) = \mathcal{N}(x \mid Bz + \mu, \sigma^2 I)$$

$$\Leftrightarrow \quad x_n := B z_n + \mu + \underbrace{\varepsilon}_{\sim \mathcal{N}(0, \sigma^2 I)}$$

$$\downarrow$$

$$p(z, x \mid B, \mu, \sigma) = p(z)\, p(x \mid z, B, \mu, \sigma)$$

$$= \mathcal{N}\left(\begin{bmatrix} z \\ x \end{bmatrix} \middle| \begin{bmatrix} 0 \\ \mu \end{bmatrix}, \begin{bmatrix} I & B^T \\ B & BB^T + \sigma^2 I \end{bmatrix} \right)$$

$B \rightarrow \mathrm{Cov}[z, x] = \mathbb{E}[zx] - \mathbb{E}[z]\mathbb{E}[x]$

(Properties of the Gaussian: Särkkä, 2011 Appendix A1)

Likelihood

$$p(x|B,\mu,\sigma) = \underbrace{\int dz\, \overbrace{p(z)}^{\mathcal{N}}\, \overbrace{p(x|z,B,\mu,\sigma)}^{\mathcal{N}}}_{\mathcal{N}}$$

$$= \mathcal{N}\left(x \mid B\cancel{0} + \mu,\ \underbrace{BIB^T}_{=BB^T} + \sigma^2 I\right)$$

PCA: $\sigma^2 \to 0$

Maximum likelihood estimation

$$\arg\max_{B,\mu,\sigma^2} \prod_{x_n} p(x_n|B,\mu,\sigma)$$

$$= \arg\max_{B,\mu,\sigma^2} \sum_{x_n} \log \mathcal{N}\left(x_n|\mu,\ \underbrace{BB^T+\sigma^2 I}_{=C}\right)$$

$$= \arg\min_{B,\mu,\sigma^2} \frac{N}{2}\left[D\log 2\pi + \log C + \mathrm{Tr}\left(C^{-1}S\right)\right]$$

$$\frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)(x_n-\mu)'$$

Posterior

$$p(z|x, B, \mu, \delta) = \mathcal{N}(z \mid m, C)$$

$$m = B^T(BB^T + \delta^2 I)^{-1}(x - \mu)$$

$$C = I - B^T(BB^T + \delta^2 I)^{-1} B$$

$= I$ when $\delta^2 \to 0$

When $\delta^2 = 0$,

$B^T(BB^T)^{-1} = B^{-1} = B^T$

$\Rightarrow m = B^T(x - \mu)$

$x = Bz + \mu$

$\Leftrightarrow z = B^T(x - \mu)$

Estimating the posterior mean

$\Leftrightarrow$ Reducing x to z

Advantages of the probabilistic perspective:

- Enable model comparison from $p(x)$
- Can use PCA to generate new data
- Provide a principled way to expand the model $\rightarrow$ ICA (non-$\mathcal{N}$ prior), Factor analysis (on element δ^2), VAE