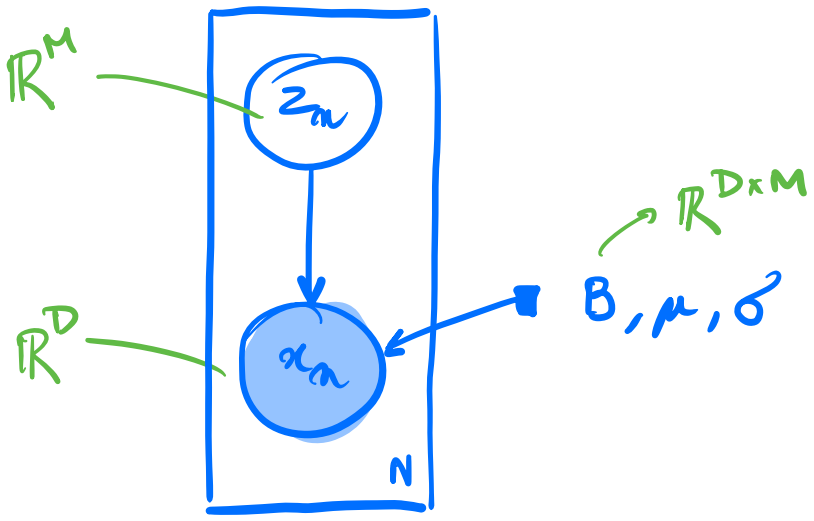
LEC 04

# Probabilistic PCA

Attention to
- min reconstruction error $\|x - BB^T x\|$
- maximum variance in $z$.




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

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

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

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

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

$$\downarrow$$

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

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

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

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

## Likelihood

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

$$= \mathcal{N}(x \mid B\cancel{0} + \mu, \; B I B^T + \sigma^2 I)$$

$= BB^T$

$\sigma^2 \to 0$ : PCA

(=) immediate by marginalization of the joint

## Maximum likelihood estimation

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

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

$$= \arg\min_{B, \mu, \sigma^2} \frac{N}{2} \Big[ d \log 2\pi + \log |C| + \underbrace{\frac{1}{2} \sum_n (x_n - \mu)^T C^{-1} (x_n - \mu)}_{= \mathrm{Tr}(SC^{-1})} \Big]$$

$$S = \sum_n (x_n - \mu)(x_n - \mu)^T$$

## Posterior

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

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

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

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

When $\sigma^2 = 0$,

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

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

$x = Bz + \mu$

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

Estimating the posterior mean

$\Leftrightarrow$ Reducing $x$ to $z$

Advantage of the probabilistic perspective:

- Enable model comparison for $p(x)$
- Can use PCA to generate new data
- Provide a principled way to expand the model → ICA (non-Gaussian prior), Factor analysis (for element $\sigma^2$), VAE