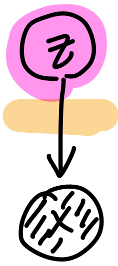
Principal component analysis



① a prior distribution over "z"

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

② a conditional dist. over "x" given "z"

$$x|z \sim N(\underbrace{Wz + b}_{\text{estimate}},\ \underbrace{\sigma^2}_{\text{fixed}} I^{d}) \in \mathbb{R}^{d}$$

$\rightarrow$ $x = Wz + b + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2 I)$

· joint dist. $p(x, z) = p(x|z)\,p(z)$

· marginal dist. $p(x) = \int_{\mathbb{R}^{d'}} p(x|z)\,p(z)\,dz = \mathbb{E}_z[p(x|z)]$

· posterior dist. $p(z|x) = \dfrac{p(x|z)\,p(z)}{p(x)} = \dfrac{p(x|z)\,p(z)}{\int_{\mathbb{R}^{d}} p(x|z)\,p(z)\,dz}$

· Marginal in PCA.

$$x = Wz + b + \varepsilon$$

$$\mathbb{E}[x] = \cancel{W\,\mathbb{E}[z]}^{\,0} + b + \cancel{\mathbb{E}[\varepsilon]}^{\,0} = b$$

$$\mathrm{Cov}[x] = \mathbb{E}[(x - \mathbb{E}[x])(x - \mathbb{E}[x])^T]$$

$$= \mathbb{E}[(Wz + \varepsilon)(Wz + \varepsilon)^T]$$

$$= WW^T + \sigma^2 I$$

~~if $\sigma^2 = 0$,~~
~~$\mathbb{E}[x] = b$~~
~~$\mathrm{Cov}[x] = WW^T$~~

· $\log p(D \mid \underbrace{W, b, \sigma^2}_{\text{implicit}}) = \sum_{n=1}^{N} \log p(x^n \mid W, b, \sigma^2)$

$$= -\frac{1}{2} \sum_{n=1}^{N} (x^n - b)^T (WW^T + \sigma^2 I)^{-1} (x^n - b)$$

$$-\frac{1}{2}\left(\log |WW^T + \sigma^2 I| + d' \log 2\pi\right)$$

$$\hat{b} = \frac{1}{N}\sum_{n=1}^{N} x^n$$

$$\frac{\partial \log |WW^T + \sigma^2 I|}{\partial W} = 2\,\mathrm{Trace}[(WW^T + \sigma^2 I)^{-1} W].$$

- Joint dist. over $[x; z]$

$$\mathbb{E}[[x; z]] = [b; 0]$$

$$\mathrm{Cov}[[x; z]] = \left[\begin{array}{c|c} \Sigma_{xx} & \Sigma_{xz} \\ \hline \Sigma_{zx} & \Sigma_{zz} \end{array}\right]$$

$WW^T + \sigma^2 I$

I

$$\Sigma_{xz} = \mathbb{E}[(x - \mathbb{E}[x])(z - \mathbb{E}[z])^T] = \mathbb{E}[(Wz + \varepsilon) z^T]$$

$$= W \underbrace{\mathbb{E}[zz^T]}_{= I} + \mathbb{E}[\varepsilon z^T] = W$$

- posterior distribution.

$$\mathbb{E}[z|x] = 0 + \Sigma_{zx}\Sigma_{xx}^{-1}(x - b) = W^T(WW^T + \sigma^2 I)^{-1}(x - b)$$

$$\mathrm{Cov}[z|x] = \Sigma_{zz} - \Sigma_{zx}\Sigma_{xx}^{-1}\Sigma_{zx} = I - W^T(WW^T + \sigma^2 I)^{-1} W$$

$$x = Wz + b + \varepsilon \rightsquigarrow x = f(z) + \varepsilon$$

the marginal $p(x)$ is not Gaussian

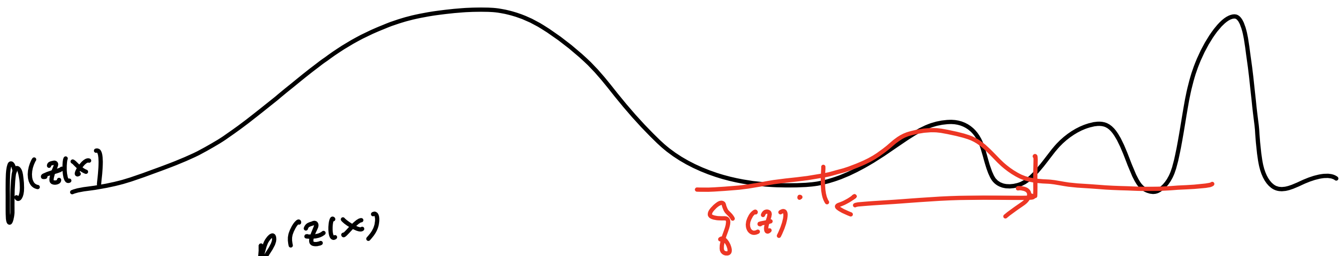
$$p(x) = \int p(x|z)\,p(z)\,dz.$$

nonlinear principal component analysis

- posterior inference $p(z|x)$
- approximate posterior inference : variational inference.

  a tractable proxy $q(z)$ (or $q(z|x)$).
- KL divergence.

$$KL(q\|p) = -\int q(z|x) \log p(z|x)\, dz + H(q).$$



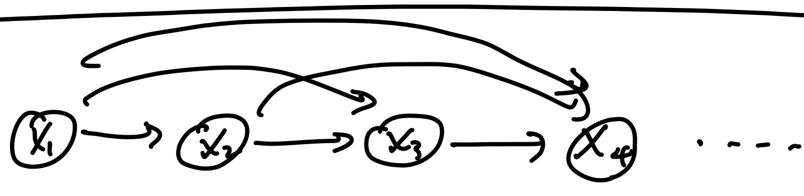
$$\underbrace{KL(q\|\overset{p(z|x)}{p})}_{\geq 0} = -\mathbb{E}_q[\log p(x|z)] + KL(q(z|x)\|p(z)) + \log p(x)$$

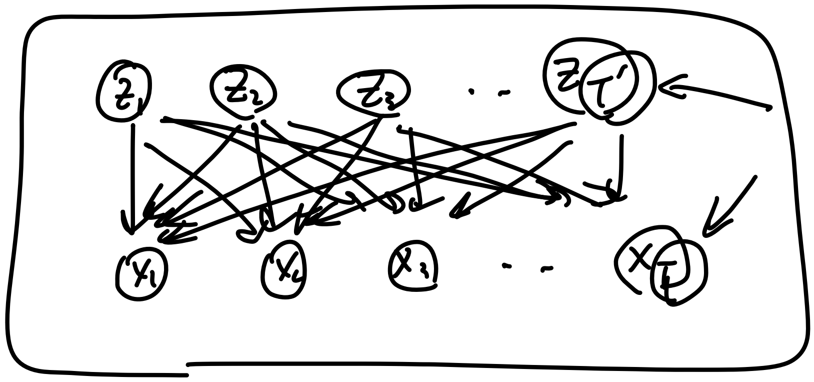
$$\log p(x) \geq \underbrace{\mathbb{E}_q[\log p(x|z)] - KL(q(z)\|p(z))}_{\text{variational lower bound}}$$

$$\boxed{\max_{q^1 \dots q^N} \max_{f} \sum_{n=1}^{N} \mathbb{E}_{q^n}[\log p(x^n|z)] - KL(q^n(z)\|p(z))}$$

$$\boxed{\begin{aligned}\max_{\theta} \max_{\phi} \sum_{n=1}^{N} \mathbb{E}_{q(z|G(x^n;\phi))}[\log p(x^n|f(z;\theta))] \\ - KL(q(z|G(x^n;\phi))\|p(z))\end{aligned}}$$

Variational autoencoder.





Latent variable sequence models