Tutorial sheet: problem 3: FA
WS15/16

= Homework sheet WS 16/17

$$p(z|\cancel{\theta}) = \mathcal{N}(\underset{L}{z} \,|\, 0, \underset{L\times L}{I})$$

$= p(z)$:-)

$$p(x|z,\theta) = \mathcal{N}(\underset{D}{x} \,|\, \underset{D\times L}{W}\underset{L}{z} + \underset{D}{\mu}, \underset{D\times D}{\Psi})$$

$\{x_i\}_{i=1}^{N}$ Training data

diagonal!

Why? We want the z to explain the correlation btw. the x.
If we allow Ψ to be a general covariance
$\Rightarrow$ Ψ explains the correlation

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

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

$$\underset{\text{(eq. 4.126 (Murphy))}}{=} \mathcal{N}(x|\mu, \Psi + WW^T)$$

* Why can we start from prior $p(z|\theta) = \mathcal{N}(z|0, I)$?

If we start from $p(z|\theta) = \mathcal{N}(z|\mu_0, \Sigma_0)$ we get:

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

$$\underset{\text{(eq 4.126 Murphy)}}{=} \mathcal{N}(x | W\mu_0 + \mu, \Psi + W\Sigma_0 W^T)$$

- absorb μ_0: We learn μ, so we can as well learn $\tilde{\mu} = W\mu_0 + \mu$ (or re-parameterize $\mu \to \mu - W\mu_0$)
- absorb Σ_0: We learn W, so we can as well learn $\tilde{W} = W\Sigma_0^{-\frac{1}{2}}$ $\left(\Sigma_0^{-1} = \Sigma_0^{-\frac{1}{2}}\Sigma_0^{-\frac{T}{2}}\right)$

$\left(\Sigma_0 = \Sigma_0^{\frac{1}{2}}\Sigma_0^{\frac{T}{2}}\right)$

e.g. via Cholesky decomp.

We then have:

$$\Psi + \tilde{W} \Sigma_0 \tilde{W}^T =$$

$$= \Psi + W \Sigma_0^{-\frac{1}{2}} \Sigma_0 \Sigma^{-\frac{T}{2}} W$$

$$= \Psi + W \Sigma_0^{-\frac{1}{2}} \Sigma_0^{\frac{1}{2}} \Sigma_0^{\frac{T}{2}} \Sigma_0^{-\frac{T}{2}} W$$

$$= \Psi + W W^T$$

Now: if we have transformed input data

$$y = Ax$$

what happens?

* General formula for linear transformations of Gaussians:

$$\boxed{\begin{array}{l} X \sim \mathcal{N}(x|\mu, \Sigma) \; ; \; Y = AX + b \\ \Rightarrow Y \sim \mathcal{N}(A\mu + b, A \Sigma A^T) \end{array}}$$

$$\Rightarrow \quad X \sim \mathcal{N}(x|\mu, \Psi + WW^T) \; ; \; Y = AX$$

$$\Rightarrow Y \sim \mathcal{N}(A\mu, A\Psi A^T + AWW^T A)$$

<u>Assume</u> the Max. likelihood solutions (e.g. won via EM) for the FA-Model for x are

$$\Theta_x^{ML} = (\mu_x, W_x, \Psi_x)$$

$\Rightarrow$ the Max. likelihood solutions for $y = Ax$ are $\Theta_y^{ML} = (A\mu_x, AW_x, A\Psi_x A^T)$

$\Rightarrow$ * case general FA: If A is diagonal: (Ψ is anyway) the characteristics of FA are preserved for y:

$$\begin{cases} A\mu_x: \text{ scaling of components of } \mu_x \\ AW_x: \text{ scaling rows of } W_x \text{ with the } a_i \ (A = \mathrm{diag}\{a_i\}) \\ A\Psi_x A^T: \text{ scaling entries in } \Psi_x \text{ with } a_i^2, \text{ still diag} \end{cases}$$

* Case 2: (PPCA) A orthogonal and $\Psi = \sigma^2 I$:

the characteristics of PPCA) are also preserved for Y:

- $A\mu_x$: Rotation of μ_x (Model is determined up to a rotation only)
- AW_x: again only a rotation applied after W → ✓
- $A\Psi_x A^T$: still diagonal → ✓

④ Tutorial sheet WS15/16 : $PPCA \xrightarrow[\sigma \to 0]{} PCA$

= Homework WS16/17

PPCA : FA with W ortho-normal ($WW^T = I$) and $\Psi = \sigma^2 I$

FA : $p(z_i | x_i) \overset{\text{(slides)}}{\underset{\text{(eq. 4.125 Murphy)}}{=}} \mathcal{N}(z_i | m_i, \Sigma)$

with $\boxed{m_i = \Sigma (W^T \Psi^{-1} (x_i - \mu)) \quad \Sigma = (I + W^T \Psi^{-1} W)^{-1}}$

$$\Sigma = (I + \sigma^{-2} W^T W)^{-1} \xleftarrow{\text{PPCA: } \Psi = \sigma^2 I}$$

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

$$\Rightarrow m_i = \sigma^2 (\sigma^2 I + W^T W)^{-1} (W^T (\sigma^2 I)^{-1} (x_i - \mu))$$

$$= (\sigma^2 I + W^T W)^{-1} (W^T (x_i - \mu))$$

__Now__: using ML solution for PPCA for W:
(closed form, no EM necessary):

$$W^{ML} = L(\Lambda - \sigma^2 I)^{\frac{1}{2}} R$$

* L: $D \times L$ matrix: (L is the dimension of the latent space)

The first L eigenvectors of the empirical covariance matrix $S = \frac{1}{N} X^T X$ for mean free $\underset{N \times D}{X}$

orthonormal because S symmetric real matrix square-root ...

$M = M^{\frac{1}{2}} M^{\frac{1}{2}}$

e.g. via Eigendecomposition

* $\Lambda = \mathrm{diag}(\lambda_1 \dots \lambda_L)$: λ_i eigenvalues of Σ.

* R : arbitrary rotation matrix

$$\Rightarrow \quad W^{ML} \underset{\sigma \to 0}{=} L \Lambda^{\frac{1}{2}} R$$

$$\Rightarrow \quad m_i \underset{\sigma \to 0}{=} (W^T W)^{-1} W^T (x_i - \mu)$$

$$= \left((L \Lambda^{\frac{1}{2}} R)^T L \Lambda^{\frac{1}{2}} R \right)^{-1} (L \Lambda^{\frac{1}{2}} R)^T (x_i - \mu)$$

$$= (R^T \Lambda^{\frac{1}{2}} \underbrace{L^T L}_{I} \Lambda^{\frac{1}{2}} R)^{-1} \; R^T \Lambda^{\frac{1}{2}} L^T (x_i - \mu)$$

$$\underset{R^{-1} = R^T}{=} R^T \Lambda^{-1} \cancel{R} \; \cancel{R^T} \Lambda^{\frac{1}{2}} L^T (x_i - \mu)$$

$$= R^T \Lambda^{-\frac{1}{2}} L^T (x_i - \mu)$$

scaling Projection

$R = I$: Same projection on L-dim subspace as in PCA