Tutoring Session 10

Unsupervised learning

Problem 1: Derive the solution $\boldsymbol{W} = \boldsymbol{U}(\Sigma - \sigma^2 \boldsymbol{I})^{\frac{1}{2}}$ of PPCA.

We start with the likelihood for PPCA, which we want to maximise.

$$\log p(\boldsymbol{X}|\boldsymbol{W}, \sigma^2) \propto \mathcal{L} = -\frac{N}{2}[\log |C| + tr(C^{-1}S)]$$

, where

$$S = \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{x}_n - \mu)(\boldsymbol{x}_n - \mu)^T$$

and

$$C = \boldsymbol{W}\boldsymbol{W}^T + \sigma^2 \boldsymbol{I}$$

We set the derivative of the likelihood to zero to find the maximum:

$$\frac{\delta \mathcal{L}}{\delta \boldsymbol{W}} = N(\boldsymbol{C}^{-1}\boldsymbol{S}\boldsymbol{C}^{-1}\boldsymbol{W} - \boldsymbol{C}^{-1}\boldsymbol{W}) \overset{!}{=} 0$$

$$\boldsymbol{C}^{-1}\boldsymbol{S}\boldsymbol{C}^{-1}\boldsymbol{W} - \boldsymbol{C}^{-1}\boldsymbol{W} = 0$$

$$\boldsymbol{C}^{-1}\boldsymbol{S}\boldsymbol{C}^{-1}\boldsymbol{W} = \boldsymbol{C}^{-1}\boldsymbol{W}$$

$$\boldsymbol{S}\boldsymbol{C}^{-1}\boldsymbol{W} = \boldsymbol{W}$$

We can solve this by setting $\boldsymbol{W} = 0$ or $\boldsymbol{S} = \boldsymbol{C}$ or finding a solution where $\boldsymbol{W} \neq 0$ and $\boldsymbol{S} \neq \boldsymbol{C}$. We will only solve for the solutions of $\boldsymbol{S} = \boldsymbol{C}$ here.

$$\boldsymbol{S} = \boldsymbol{C}$$

$$\boldsymbol{S} = \boldsymbol{W}\boldsymbol{W}^T + \sigma^2 \boldsymbol{I}$$

$$\boldsymbol{W}\boldsymbol{W}^T = \boldsymbol{S} - \sigma^2 \boldsymbol{I}$$

We decompose the positive S into $\boldsymbol{U}\Sigma\boldsymbol{U}^T$ with the singular value decomposition.

$$\boldsymbol{W}\boldsymbol{W}^T = \boldsymbol{U}\Sigma\boldsymbol{U}^T - \sigma^2\boldsymbol{I}$$

$$\boldsymbol{W}\boldsymbol{W}^T = \boldsymbol{U}[\Sigma - \boldsymbol{U}^T\sigma^2\boldsymbol{I}\boldsymbol{U}]\boldsymbol{U}^T$$

$$\boldsymbol{W}\boldsymbol{W}^T = \boldsymbol{U}[\Sigma - \sigma^2\boldsymbol{I}]\boldsymbol{U}^T$$

$$\boldsymbol{W}\boldsymbol{W}^T = \boldsymbol{U}[\Sigma - \sigma^2\boldsymbol{I}]^{\frac{1}{2}}[\Sigma - \sigma^2\boldsymbol{I}]^{\frac{1}{2}}\boldsymbol{U}^T$$

$$\boldsymbol{W}\boldsymbol{W}^T = \boldsymbol{U}[\Sigma - \sigma^2\boldsymbol{I}]^{\frac{1}{2}}[\Sigma - \sigma^2\boldsymbol{I}]^{\frac{1}{2}^T}\boldsymbol{U}^T$$

$$\boldsymbol{W}\boldsymbol{W}^T = (\boldsymbol{U}[\Sigma - \sigma^2\boldsymbol{I}]^{\frac{1}{2}})(\boldsymbol{U}[\Sigma - \sigma^2\boldsymbol{I}]^{\frac{1}{2}})^T$$

$$\boldsymbol{W} = \boldsymbol{U}[\Sigma - \sigma^2\boldsymbol{I}]^{\frac{1}{2}}$$

Problem 2: Imputation in a Factor analysis model: Derive the expression for $p(x_h|x_v,\theta)$ for a FA model. x_v contains all visible data-dimensions while x_h contains all hidden ones.

From the MVN slides we know:

$$p(x) = p(x_I|x_R)p(x_R)$$
$$\mu_{I|R} = \mu_I + \Sigma_{IR}\Sigma_{RR}^{-1}(x_R - \mu_R)$$
$$\Sigma_{I|R} = \Sigma_{II} - \Sigma_{IR}\Sigma_{RR}^{-1}\Sigma_{RI}$$

for FA we get:

$$C = WW^T + \Phi$$

$$p(x_h|x_v,\theta) = \mathcal{N}(x_h|\mu_h + C_{hv}C_{vv}^{-1}(x_v - \mu_v), C_{hh} - C_{hv}C_{vv}^{-1}C_{vh})$$