Machine Learning Worksheet 10

Unsupervised Learning

Problem 1: Where would you use PCA and where ICA on real world data?

Problem 2: Consider the latent space distribution

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

and a conditional distribution for the observed variable $\boldsymbol{x} \in \mathbb{R}^d$,

$$p(\boldsymbol{x}|\boldsymbol{z}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{W}\boldsymbol{z} + \boldsymbol{\mu}, \boldsymbol{\Phi})$$

where $\boldsymbol{\Phi}$ is an arbitrary symmetric, positive-definite noise covariance variable. Furthermore, $\boldsymbol{A}$ is a non-singular $d \times d$ matrix and $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$. Show that for the maximum likelihood solution for the parameters of the model for $\boldsymbol{y}$ specific constraints on $\boldsymbol{\Phi}$ are preserved in the following two cases: (i) $\boldsymbol{A}$ is a diagonal matrix and $\boldsymbol{\Phi}$ is a diagonal matrix (this corresponds to the case of Factor Analysis). (ii) $\boldsymbol{A}$ is orthogonal and $\boldsymbol{\Phi} = \sigma^2\boldsymbol{I}$ (this corresponds to pPCA).

Problem 3: Show that in the limit $\sigma^2 \to 0$ the posterior mean for the probabilistic PCA model becomes an orthogonal projection onto the same principal subspace as in PCA.

You may use the solution for the posterior of the FA model:

$$p(\boldsymbol{z}_i|\boldsymbol{x}_i) = \mathcal{N}(\boldsymbol{z}_i|\boldsymbol{m}_i, \boldsymbol{\Sigma})$$

$$\boldsymbol{\Sigma} = (\boldsymbol{I} + \boldsymbol{W}^T\boldsymbol{\Psi}^{-1}\boldsymbol{W})^{-1}$$

$$\boldsymbol{m}_i = \boldsymbol{\Sigma}(\boldsymbol{W}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}))$$