Exercises for the course

Machine Learning 2

Summer semester 2019

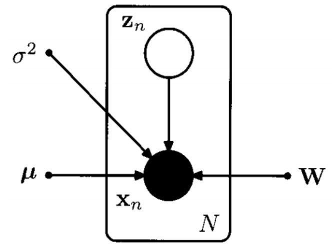
Abteilung Maschinelles Lernen
Institut für Softwaretechnik und theoretische Informatik
Fakultät IV, Technische Universität Berlin
Prof. Dr. Klaus-Robert Müller
Email: klaus-robert.mueller@tu-berlin.de

Exercise Sheet 5

Recall: Let $x \in \mathbb{R}^d$ be an input vector drawn from some distribution $p(x)$. The probabilistic PCA model is defined by an explicit latent variable $z \in \mathbb{R}^m$ so $m < d$, which corresponds to a principal component subspace. The prior distribution of the latent variable z is given by $p(z) = \mathcal{N}(0, \mathbb{I}_m)$. The distribution of the input vector x conditioned on the variable z and all the model parameters $\theta = \{W, \mu, \sigma^2\}$ has then the following form:

$$p(x|z, \theta) = \mathcal{N}(Wz + \mu, \sigma^2 \mathbb{I}_d),$$

where W is an expansion matrix of size $d \times m$, $\mu \in \mathbb{R}^d$ is a mean and $\sigma^2 \in \mathbb{R}^+$ is a variance parameter. The corresponding directed graph for the probabilistic PCA model can be depicted as following:



We would like to derive the EM algorithm for probabilistic PCA, that allows to find the model parameters that best fit the data. In order to do it, we have to follow a general framework for the EM algorithm.

Exercise 1: Complete-data log likelihood function (25 P)

First we need to write down the complete-data log likelihood. Let $X = (x_n)_{n=1}^N$ be a collection of data points drawn i.i.d., and $Z = (z_n)_{n=1}^N$ be the corresponding latent states. For i.i.d. data, the complete-data log likelihood function can generally be written as:

$$\log p(X, Z \mid \theta) = \sum_{n=1}^{N} \{\log p(x_n|z_n; \theta) + \log p(z_n|\theta)\} \tag{1}$$

where θ denotes the parameters of the model.

(a) *Give* the explicit form of Eq. (1) for probabilistic PCA, by making use of the latent and conditional distributions defined above. That is, write the function $\log p(X, Z \mid W, \mu, \sigma^2)$.

Note: You can write the log of the d-dimensional Gaussian distribution using the trace operator as following:

$$\ln \mathcal{N}_d(x|\mu, \Sigma) = -\frac{1}{2}\Big(d \ln 2\pi + \ln|\Sigma| + \operatorname{tr}\{\Sigma^{-1}(x-\mu)(x-\mu)^\top\}\Big) \tag{2}$$

Exercise 2: E-Step updates (50 P)

(a) *Show* that the E-step for the latent variable z_n can be expressed as:

$$\mathbb{E}[z_n|x_n] = (W^\top W + \sigma^2 \mathbb{I}_m)^{-1} W^\top (x_n - \mu)$$

and

$$\mathbb{E}[z_n z_n^\top|x_n] = \sigma^2 (W^\top W + \sigma^2 \mathbb{I}_m)^{-1} + \mathbb{E}[z_n|x_n] \cdot \mathbb{E}[z_n|x_n]^\top$$

Note: In this exercise, you can make use of the completing squares trick or the usual conditioning identities for multivariate Gaussian distributions (c.f. the Matrix Cookbook).

Exercise 3: M-Step updates (25 P)

The M-step can be derived from maximizing complete-data log likelihood, while considering the expectations computed in the E-step as ground-truth.

(a) *Compute* for the variance parameter σ^2 its updated value σ^2_{new} after observing N data points.