UIBK

# Deep Learning

Homework №3, task 1

**Student:** Nikulina Kseniia

Innsbruck, 2022

# 1 EM solution for Probabilistic PCA

Probabilistic PCA represents a constrained form of the Gaussian distribution in which the number of free parameters can be restricted while still allowing the model to capture the dominant correlations in a data set.

## 1.1 Parameters that are learned

Computed the sufficient statistics of the latent space posterior distribution and then revised the parameter values

## 1.2 E-step equations

Write down the complete-data log likelihood and take its expectation with respect to the posterior distribution of the latent distribution evaluated using 'old' parameter values.
The complete-data log likelihood function takes the form

$$\ln p\left(\mathbf{X}, \mathbf{Z} | \boldsymbol{\mu}, \mathbf{W}, \sigma^2\right) = \sum_{n=1}^{N} \{\ln p(\mathbf{x}_n | \mathbf{z}_n) + \ln p(\mathbf{z}_n)\}$$

Making use of the expressions

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

and

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

for the latent and conditional distributions, respectively, and taking the expectation with respect to the posterior distribution over the latent variables, we obtain

$$\begin{aligned}\mathbb{E}[\ln p\left(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \mathbf{W}, \sigma^2\right)] = -\sum_{n=1}^{N}\Bigg\{&\frac{D}{2}\ln(2\pi\sigma^2) + \frac{1}{2}\mathrm{Tr}\left(\mathbb{E}[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}]\right) \\ &+\frac{1}{2\sigma^2}\|\mathbf{x}_n - \boldsymbol{\mu}\|^2 - \frac{1}{\sigma^2}\mathbb{E}[\mathbf{z}_n]^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}(\mathbf{x}_n - \boldsymbol{\mu}) \\ &+ \frac{1}{2\sigma^2}\mathrm{Tr}\left(\mathbb{E}[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}]\mathbf{W}^{\mathrm{T}}\mathbf{W}\right)\Bigg\}.\end{aligned}$$

We use the old parameter values to evaluate

$$\mathbb{E}[z_n] = M^{-1}W^T(x_n - \bar{x})$$

$$\mathbb{E}[z_n z_n^T] = \sigma^2 M^{-1} + \mathbb{E}[z_n]\mathbb{E}[z_n]^T$$

Which follow directly from the posterior distribution

$$p(z|x) = \mathcal{N}(z|M^{-1}W^T(x - \mu), \sigma^(-2)M)$$

together with the standard result

$$\mathbb{E}[z_n z_n^T] = cov[z_n] + \mathbb{E}[z_n]\mathbb{E}[z_n]^T$$

where M is

$$M = W^T W + \sigma^2 I$$

### 1.3 M-step equations

In the M step, we maximize with respect to W and $\sigma^2$, keeping the posterior statistics fixed. Maximization with respect to $\sigma^2$ is straightforward. For the maximization with respect to W we obtain the M-step equations

$$\begin{aligned} \mathbf{W}_{\text{new}} &= \left[\sum_{n=1}^{N}(\mathbf{x}_n - \overline{\mathbf{x}})\mathbb{E}[\mathbf{z}_n]^{\mathrm{T}}\right]\left[\sum_{n=1}^{N}\mathbb{E}[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}]\right]^{-1} \\ \sigma^2_{\text{new}} &= \frac{1}{ND}\sum_{n=1}^{N}\left\{\|\mathbf{x}_n - \overline{\mathbf{x}}\|^2 - 2\mathbb{E}[\mathbf{z}_n]^{\mathrm{T}}\mathbf{W}_{\text{new}}^{\mathrm{T}}(\mathbf{x}_n - \overline{\mathbf{x}}) \right. \\ & \left. +\mathrm{Tr}\left(\mathbb{E}[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}]\mathbf{W}_{\text{new}}^{\mathrm{T}}\mathbf{W}_{\text{new}}\right)\right\}. \end{aligned}$$

### 1.4 EM algorithm for Probabilistic PCA

Iterative process to estimate parameters consisting of two steps for each iteration.

**Expectation (data step):** proceeds by initializing the parameters and then alternately computing the sufficient statistics of the latent space posterior distribution using $\mathbb{E}[z_n]$ and $\mathbb{E}[z_n z_n^T]$

- Fix subspace and project data, y, into it to give values of hidden states x
- Known: Y: d-dimensional observed data
- Unknown (latent): X: q-dimensional unknown states

**Maximization (likelihood step):** Update set of parameters, using using $W_{new}$ and $\sigma^2_{new}$, from complete set of data from previous step

- Fix values of hidden states and choose subspace orientation that minimizes squared reconstruction errors