Übung zur Vorlesung

Maschinelles Lernen 2

Sommersemester 2013

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 3

Due May 6, 2013, at 10:00 a.m. local time in electronic form via the ISIS website. See the corresponding submission page for formatting details (.zip archives, source code without fancy non-ASCII symbols etc.)

3. **The SSA Cost Function (30 + 20 P)**

(a) Let $X_1 \sim \mathcal{N}(\mu_1, \Sigma_1)$ and $X_2 \sim \mathcal{N}(\mu_2, \Sigma_2)$ be Gaussian random variables with values in $\mathbb{R}^n$, with mean $\mu_i \in \mathbb{R}^n$ and covariance matrix $\Sigma_i \in \mathbb{R}^{n \times n}$, i.e., the probability density function of X_i is

$$p_i(x) = ((2\pi)^n \det \Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu_i)^\top \Sigma^{-1}(x-\mu_i)\right).$$

Derive the following explicit formula for the KL-divergence from X_2 to X_1:

$$D_{\mathrm{KL}}\left(X_2 \mid\mid X_1\right) = \frac{1}{2}\left(\log\left(\frac{\det \Sigma_1}{\det \Sigma_2}\right) + \operatorname{tr}\left(\Sigma_1^{-1}\Sigma_2\right) + (\mu_1 - \mu_2)^\top \Sigma_1^{-1}(\mu_1 - \mu_2) - n\right).$$

(b) *Derive* the cost function for gradient based SSA. That is, consider the following scenario:

Let D be the dimension of the input data, $d < D$ the number of stationary sources and N the number of epochs. Denote the estimates of the means and covariances in each of the N epochs by $\hat{\boldsymbol{\mu}}_1, \ldots, \hat{\boldsymbol{\mu}}_N \in \mathbb{R}^D$ and $\hat{\Sigma}_1, \ldots, \hat{\Sigma}_N \in \mathbb{R}^{D \times D}$. In SSA, a centering and whitening is performed on the data; that is, you may assume that

$$\sum_{i=1}^N \hat{\boldsymbol{\mu}}_i = 0 \quad \text{and} \quad \sum_{i=1}^N \hat{\Sigma}_i = I,$$

where I is the $(D \times D)$ identity matrix. Given the rotational component $R \in \mathbb{R}^{D \times D}$ (with $R^\top R = I$) of the estimated demixing matrix, the mean and covariance matrix of the estimated stationary sources of epoch i can be formulated as

$$\hat{\boldsymbol{\mu}}_i^{\mathfrak{s}} = I^d R \hat{\boldsymbol{\mu}}_i \quad \text{and} \quad \hat{\Sigma}_i^{\mathfrak{s}} = I^d R \hat{\Sigma}_i (I^d R)^\top,$$

where $I^d \in \mathbb{R}^{d \times D}$ is the identity matrix, truncated to the first d rows.

Show the following explicit formula for the SSA cost function:

$$\begin{aligned} L(R) &= \sum_{i=1}^N D_{\mathrm{KL}}\left[\mathcal{N}(\hat{\boldsymbol{\mu}}_i^{\mathfrak{s}}, \hat{\Sigma}_i^{\mathfrak{s}}) \mid\mid \mathcal{N}(0, I)\right] \\ &= \frac{1}{2}\sum_{i=1}^N \left(-\log\det \hat{\Sigma}_i^{\mathfrak{s}} + \hat{\boldsymbol{\mu}}_i^{\mathfrak{s}\top} \hat{\boldsymbol{\mu}}_i^{\mathfrak{s}}\right) - \frac{N-1}{2} d. \end{aligned}$$

This means: *show* that the first line, written out, equates to the second.

Hint: Use the explicit formula for the Kullback-Leibler divergence derived in (a).

4. **Finding the stationary subspace (50 P)**
On ISIS, you can find a data set and a MATLAB implementation of SSA. *Determine* the number of stationary sources, and the projection onto the stationary sources. *Plot* the unmixed sources. *Write* a brief report on what you have done. That is, *describe* the data set, which experiments you have done, *make* suitable plots, and *discuss* your findings.

5. **Uniqueness and identifiability of the SSA model (10* + 10* + 10* P)**
Assume the standard mixing model for SSA, i.e.,

$$\mathbf{x}(t) = A\mathbf{s}(t) = \begin{bmatrix} A^{\mathfrak{s}} & A^{\mathfrak{n}} \end{bmatrix} \begin{bmatrix} \mathbf{s}^{\mathfrak{s}}(t) \\ \mathbf{s}^{\mathfrak{n}}(t) \end{bmatrix}, \tag{1}$$

where $\mathbf{s}^{\mathfrak{s}}(t) = [s_1(t), s_2(t), \ldots, s_d(t)]^\top$ are the stationary sources,
$\mathbf{s}^{\mathfrak{n}}(t) = [s_{d+1}(t), s_{d+2}(t), \ldots, s_D(t)]^\top$ are non-stationary sources, and $A \in \mathbb{R}^{D\times D}$ is the mixing matrix. Furthermore, assume that the number of stationary sources d is exactly known.

(a*) *Show*, by giving an example: Given any data sample $\mathbf{x}(t)$, the mixing matrix A can in general not be identified exactly.

Furthermore, *prove* that in every model for A which only assumes that $\mathbf{s}^{\mathfrak{s}}(t)$ is stationary and $\mathbf{s}^{\mathfrak{n}}(t)$ is non-stationary, the matrix $A^{\mathfrak{n}} \in \mathbb{R}^{D\times(D-d)}$ is only determined up to right multiplication with an invertible $((D-d)\times(D-d))$ matrix, and $A^{\mathfrak{s}}$ is not determined at all (from perfect knowledge of all possible data).

(b*) *Show* that the mixing model in Equation 1 is uniquely determined by the column span of $A^{\mathfrak{n}}$. That is, show that the invariances found in exercise (a) are all there exist.

(c*) Assume $D = 2, d = 1$. Furthermore assume that exact epoch covariance matrices $\Sigma_1, \ldots, \Sigma_N$ are available. In the cases $N = 1, 2, 3, 4$, *discuss* whether the Σ_i suffice to determine $A^{\mathfrak{s}}$ uniquely, and/or under which conditions.

Please ask questions in the ISIS discussion forums for Machine Learning 2:
`https://www.isis.tu-berlin.de/course/view.php?id=8005`