Machine learning 3
Exercise sheet 3

FLEISCHMANN Kay, Matrnr: 352247
ROHRMANN Till, Matrnr: 343756

April 29, 2013

3 The SSA Cost Function

Given $X_1 \sim \mathcal{N}(\mu_1, \Sigma_1)$ and $X_2 \sim \mathcal{N}(\mu_2, \Sigma_2)$ be Gaussian random variables with values in $\mathbb{R}^n$. The probability density function of X_i is

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

Derive the explicit formula for the KL-divergence between X_1 and X_2:

$$\begin{aligned}
D_{KL}(X_2 \,||\, X_1) &= \int p_2(x) \log\left(\frac{p_2(x)}{p_1(x)}\right)\, dx && (2)\\
&= \int p_2(x) \left(-\frac{1}{2}\log\left((2\pi)^n \det \Sigma_2\right) - \frac{1}{2}(x-\mu_2)^T \Sigma_2^{-1}(x-\mu_2)\right.\\
&\quad \left. + \frac{1}{2}\log\left((2\pi)^n \det \Sigma_1\right) + \frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1)\right)\, dx && (3)\\
&= \frac{1}{2}\log\left(\frac{\det \Sigma_1}{\det \Sigma_2}\right) + \int p_2(x)\left(-\frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)\right.\\
&\quad \left. + \frac{1}{2}(x-\mu_2+\mu_2-\mu_1)^T \Sigma_1^{-1}(x-\mu_2+\mu_2-\mu_1)\right)\, dx && (4)\\
&= \frac{1}{2}\left(\log\left(\frac{\det \Sigma_1}{\det \Sigma_2}\right) - \mathbb{E}\left[(x-\mu_2)^T \Sigma_2^{-1}(x-\mu_2)\right]\right.\\
&\quad + \mathbb{E}\left[(x-\mu_2)^T\Sigma_1^{-1}(x-\mu_2)\right] + \mathbb{E}\left[(x-\mu_2)^T\Sigma_1^{-1}(\mu_2-\mu_1)\right]\\
&\quad \left. \mathbb{E}\left[(\mu_2-\mu_1)^T\Sigma_1^{-1}(x-\mu_2)\right] + (\mu_2-\mu_1)^T\Sigma_1^{-1}(\mu_2-\mu_1)\right) && (5)
\end{aligned}$$

Assume ϵ is a vector of n random variables. Let μ denote its mean and Σ its covariance matrix. Let Λ be an n-dimensional symmetric matrix.

$$\mathbb{E}\left[\epsilon^T \Lambda \epsilon\right] = \mathrm{tr}\left(\mathbb{E}\left[\epsilon^T \Lambda \epsilon\right]\right) \tag{6}$$

Due to the linearity of tr it holds $\mathbb{E} \circ \mathrm{tr} = \mathrm{tr} \circ \mathbb{E}$ and thus

$$\mathbb{E}\left[\epsilon^T \Lambda \epsilon\right] = \mathbb{E}\left[\mathrm{tr}\left(\epsilon^T \Lambda \epsilon\right)\right] \tag{7}$$

Due to the circular property of the trace operator $\mathrm{tr}(ABC) = \mathrm{tr}(BCA)$ the following equation holds

$$\begin{aligned}
\mathbb{E}\left[\epsilon^T \Lambda \epsilon\right] &= \mathbb{E}\left[\mathrm{tr}\left(\Lambda \epsilon \epsilon^T\right)\right] && (8)\\
&= \mathrm{tr}\left(\Lambda \mathbb{E}\left[\epsilon\epsilon^T\right]\right) && (9)\\
&= \mathrm{tr}\left(\Lambda\left(\Sigma + \mu\mu^T\right)\right) && (10)\\
&= \mathrm{tr}\left(\Lambda\Sigma\right) + \mu^T \Lambda \mu && (11)
\end{aligned}$$

Applied to equation (5) we obtain

$$\begin{aligned}
(5) &= \frac{1}{2}\left(\log\left(\frac{\det \Sigma_1}{\det \Sigma_2}\right) - \mathrm{tr}\left(\Sigma_2^{-1}\Sigma_2\right) + \mathrm{tr}\left(\Sigma_1^{-1}\Sigma_2\right) + 0 + 0\right.\\
&\quad \left. + (\mu_2-\mu_1)^T\Sigma_1^{-1}(\mu_2-\mu_1)\right) && (12)\\
&= \frac{1}{2}\left(\log\left(\frac{\det \Sigma_1}{\det \Sigma_2}\right) - n + \mathrm{tr}\left(\Sigma_1^{-1}\Sigma_2\right) + (\mu_2-\mu_1)^T\Sigma_1^{-1}(\mu_2-\mu_1)\right) && (13)
\end{aligned}$$

Which is the final result.
Show that the explicit formula for the SSA cost function can be written:

$$L(R) = \sum_{i=1}^{N} D_{KL}\left[\mathcal{N}(\hat{\mu}_i^s, \hat{\Sigma}_i^s) \,||\, \mathcal{N}(0, I)\right] \tag{14}$$

$$= \frac{1}{2}\sum_{i=1}^{N}\left(-\log\left(\det \hat{\Sigma}_i^s\right) + (\hat{\mu}_i^s)^T \hat{\mu}_i^s\right) - \frac{N-1}{2}d \tag{15}$$

Proof.

$$D_{KL}\left[\mathcal{N}(\hat{\mu}_i^s, \hat{\Sigma}_i^s) \,||\, \mathcal{N}(0, I)\right] = \frac{1}{2}\left(\log\left(\frac{1}{\det \hat{\Sigma}_i^s}\right) + \mathrm{tr}(\hat{\Sigma}_i^s) + (\hat{\mu}_i^s)^T(\hat{\mu}_i^s) - d\right) \tag{16}$$

$$\sum_{i=1}^{N} D_{KL}\left[\mathcal{N}(\hat{\mu}_i^s, \hat{\Sigma}_i^s) \,||\, \mathcal{N}(0, I)\right] = \frac{1}{2}\sum_{i=1}^{N}\left(-\log\left(\det \hat{\Sigma}_i^s\right) + (\hat{\mu}_i^s)^T(\hat{\mu}_i^s)\right) + \frac{1}{2}\sum_{i=1}^{N}\mathrm{tr}(\hat{\Sigma}_i^s) - \frac{N}{2}d \tag{17}$$

$$\sum_{i=1}^{N}\mathrm{tr}(\hat{\Sigma}_i^s) = \mathrm{tr}\left(\sum_{i=1}^{N}\hat{\Sigma}_i^s\right) \tag{18}$$

$$= \mathrm{tr}\left(\sum_{i=1}^{N} I^d R \hat{\Sigma}_i (I^d R)^T\right) \tag{19}$$

$$= \mathrm{tr}\left(I^d R\left(\sum_{i=1}^{N}\hat{\Sigma}_i\right)(I^d R)^T\right) \tag{20}$$

$$= \mathrm{tr}\left(I^d R I (I^d R)^T\right) \tag{21}$$

$$= d \tag{22}$$

By inserting this result into equation (17) we finally obtain:

$$L(R) = \frac{1}{2}\sum_{i=1}^{N}\left(-\log\left(\det \hat{\Sigma}_i^s\right) + (\hat{\mu}_i^s)^T(\hat{\mu}_i^s)\right) - \frac{N-1}{2}d \tag{23}$$

□

4 Finding the stationary subspace