Exercise sheet 10 To be handed in on **Thursday, 30.06.2016**

Principal Component Analysis

1 Group exercises

G 1. Let $\Sigma \in \mathbb{R}^{d\times d}$ be a symmetric matrix with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_d \geq 0$.

a) Let $u_i \in \mathbb{R}^d$ be the eigenvector corresponding to eigenvalue λ_i. Show that $\langle u_i, u_j\rangle = 0$ if $\lambda_i \neq \lambda_j$.

b) Show that

$$\max_{\|w\|_2=1} w^T\Sigma w = \lambda_1, \qquad \min_{\|w\|_2=1} w^T\Sigma w = \lambda_d.$$

G 2. Let $\lambda_1, \ldots, \lambda_d \in \mathbb{R}$ such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d \geq 0$. Further let $\alpha_1, \ldots, \alpha_d \in [0, 1]$ such that $\sum_{i=1}^d \alpha_i = p$, where $p \in \mathbb{N}$ and $p \leq d$. Show that

$$\sum_{i=1}^{d} \lambda_i\alpha_i \leq \sum_{i=1}^{p} \lambda_i.$$

G 3. Let $y \in \mathbb{R}^d$ be a random variable with $\mathbb{E}[y] = 0$ and $\mathbb{E}[yy^T] = \Sigma_y$. Show that the minimization problem

$$\min_{W\in\mathbb{R}^{d\times p}: W^TW=I_p} \mathbb{E}[\|y - WW^Ty\|_2^2]$$

is equivalent to the maximization problem

$$\max_{W\in\mathbb{R}^{d\times p}: W^TW=I_p} \operatorname{tr}(W^T\Sigma_y W)$$

G 4. Let $y^1, \ldots, y^n$ be some data points in $\mathbb{R}^d$. Assume that you cannot access the y^i directly, but only the Euclidean distance matrix $D = (\|y^i - y^j\|_2^2)_{i,j=1}^n$. Compute from D the centering Gram matrix $G^c = (\langle y^i - \mu, y^j - \mu\rangle)_{i,j=1}^n$, where $\mu = n^{-1}\sum_{i=1}^n y^i$.

2 Homework

H 1. (Principal components)

5 Let $y \in \mathbb{R}^d$ be a random variable with $\mathbb{E}[y] = 0$ and $\Sigma_y = \mathbb{E}[yy^T]$. Assume that $\operatorname{rank}(\Sigma_y) \geq p$ and denote by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$ the p largest eigenvalues of Σ_y. Prove Theorem 2.3 presented in the lecture, i.e., show that the first p principal components of y are given by

$$x_i = w_i^T y$$

where $\{w_i\}_{i=1}^p$ are the orthonormal eigenvectors of Σ_y associated with the eigenvalues $\lambda_1, \ldots, \lambda_p$.
Hint: Use G1 b) to determine the p principal components step by step.

H 2. (PCA: Greedy vs. global minimization)

a) A *greedy algorithm* is an algorithm that tries to solve an optimization problem by making a locally optimal choice at each stage. A solution obtained by a greedy algorithm is called a *greedy solution.* Argue that the principal components, which you determined in H1, form a greedy solution for the maximiation problem considered in G3.

b) Let $\Sigma \in \mathbb{R}^{d\times d}$ be a symmetric matrix with eigenvalue decomposition $\Sigma = U\Lambda U^T$. In the lecture it has been stated that

$$\max_{W\in\mathbb{R}^{d\times p}, W^TW=I_p} \operatorname{tr}(W^T\Sigma W)$$

is attained for $W = UI_{d\times p}$, where $I_{d\times p} = (\delta_{ij})_{i=1,\ldots,d,j=1,\ldots,p}$. Show that this is indeed true. **Hint:** Use G2.

c) Use b) to conclude the principal components from H1 give the optimal solution for the optimization problem considered in G3.

Remark: This fact is remarkable since in general, a greedy algorithm will not find the optimal solution of an optimization problem.

(5 Punkte)

H 3. (PCA: optimal rank-p approximation)

Let $A \in \mathbb{R}^{n\times d}$ with singular value decomposition $A = U\Gamma V^T$. Let $A^p = U\Gamma^P V^T$, where Γ^p denotes the matrix obtained from Γ by settings to zero its elements on the diagonal after the pth entry.

a) Show that

$$\|A - A^p\|_F^2 = \sum_{i=p+1}^{\min\{n,d\}} \sigma_i^2,$$

where σ_i denotes the ith singular value of A.

b) Show that the solution of

$$\min_{B\in\mathbb{R}^{n\times d}, \operatorname{rank}(B)=p} \|A - B\|_F^2$$

is given by $B = A^p$.

(5 Punkte)

H 4. (Programming exercise: oddities of high-dimensional data)

In this programming exercise you will experiment with artificial high-dimensional data to observe that distances behave very counterintuitive in high dimensions. See accompanying notebook for the details.

(5 Punkte)