ML1 (WS 2016/17)
Exercise Sheet 1

Group SWVTI

(Gaurav Vashish, Nhat Tran, René Saitenmacher, Sebastian Walther, Matthias Ivantsits)

Excercise 1:

The objective function without constraints is given by

$$J(\theta) = \sum_{k=1}^{n} \|\theta - x_k\|^2$$

(a)

Now, a constraint is given by $c_1(\theta, b) = \theta^T b$. The associated Langrange Multiplier therefore is

$$\Lambda(x, \theta, b) = \sum_{k=1}^{n} \|\theta - x_k\|^2 + \lambda(\theta^T b)$$

Generall, the euclidian distance between two vectors is defined by

$$\|a - b\| = \left(\sum_{j=1}^{d} (a-b)^2 \right)^{0.5}$$

Therefore, the derivate to vector a is given by

$$\frac{\delta(\|a-b\|)}{\delta a} = \begin{pmatrix} 0.5 * \left(\sum_{j=1}^{d} (a-b)^2 \right)^{-0.5} * 2 * (a_1 - b_1) \\ . \\ . \\ . \\ 0.5 * \left(\sum_{j=1}^{d} (a-b)^2 \right)^{-0.5} * 2 * (a_d - b_d) \end{pmatrix}$$

$$= \begin{pmatrix} \frac{a_1 - b_1}{\|a-b\|} \\ . \\ . \\ . \\ \frac{a_d - b_d}{\|a-b\|} \end{pmatrix}$$

$$= \frac{1}{\|a-b\|} \begin{pmatrix} a_1 - b_1 \\ . \\ . \\ . \\ a_d - b_d \end{pmatrix}$$

$$= \frac{a-b}{\|a-b\|}$$

Back to the Lagrange Multiplier. In order to find the minimum, we set the first derivate to zero:

$$\begin{aligned}\frac{\delta\Lambda(x,\theta,b)}{\delta\theta} &= 2*\sum_{k=1}^{n}\|\theta - x_k\| * \|\theta - x_k\|' + \lambda b \stackrel{!}{=} 0 \\ \Rightarrow 2*n*\|\theta-\bar{x}\| * \|\theta-\bar{x}\|' &= -\lambda b \\ \Rightarrow \|\theta-\bar{x}\| * \frac{\theta-\bar{x}}{\|\theta-\bar{x}\|} &= -\frac{\lambda}{2n}b \\ \Rightarrow \theta^* &= -\frac{\lambda}{2n}b + \bar{x}\end{aligned}$$

(b)

Now, a constraint is given by $c_2(\theta, c) = \|\theta - c\|^2 = 1$. The associated Langrange Multiplier therefore is

$$\Lambda(x,\theta,b) = \sum_{k=1}^{n}\|\theta - x_k\|^2 + \lambda(\|\theta - c\|^2 - 1)$$

In order to find the minimum, we set the first derivate to zero:

$$\begin{aligned}\frac{\delta\Lambda(x,\theta,c)}{\delta\theta} &= 2n(\theta-\bar{x}) + 2\lambda(\theta - c) \stackrel{!}{=} 0 \\ \Rightarrow \theta(n+\lambda) &= n\bar{x} + \lambda c \\ \Rightarrow \theta^* &= \frac{n\bar{x}+\lambda c}{n+\lambda}\end{aligned}$$

Excercise 2:

(a)

Note that

$$\sum_{i=1}^{d} S_{ii} = tr(S)$$

Also, S can be decomposed in

$$S = W\Lambda W^T$$

where $WW^T = I$, e.g. W is orthogonal.
Generally, it holds for an invertible matrix B that

$$tr(B^{-1}AB) = tr(A)$$

Therefore

$$tr(S) = tr(W\Lambda W^T) = tr(\Lambda)$$

Since $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d \geq 0$:

$$\lambda_1 \leq tr(\Lambda) = tr(S)$$

(b)

Since $tr(S) = tr(\Lambda)$, the upper bound is tight iff $tr(\Lambda) = \lambda_1$ which means that λ_1 is the only non-zero eigenvalue, i.e. the data only variates along one dimension.

(c)

Let $S_{jj} = \max_{i \in [d]} S_{ii}$ for some $j \in [d]$ and let $y \in R^d$ such that $y_j = 1$ and $\forall i \in [d], i \neq j : y_i = 0$.

$$
\begin{aligned}
& \lambda_1 \overset{!}{\geq} S_{jj} \\
\iff \quad & \lambda_1 \overset{!}{\geq} y S y^T \\
\iff \quad & \lambda_1 \overset{!}{\geq} y W \Lambda W^T y^T \\
\iff \quad & \lambda_1 \overset{!}{\geq} y W \Lambda (yW)^T \\
\underset{(1)}{\iff} \quad & \lambda_1 \overset{!}{\geq} \sum_{i=1}^{d} \lambda_i (yW)_i^2
\end{aligned}
$$

(1): Λ is diagonal matrix

But, one can make the following observation:

$$\sum_{i=1}^{d} (yW)_i^2 = yW(yW)^T = yWW^Ty^T = yIy^T = yy^T = 1$$

Using this observation and the fact that $\forall i \in [d] : \lambda_1 \geq \lambda_i$, we can conclude:

$$\lambda_1 = \sum_{i=1}^{d} \lambda_1 (yW)_i^2 \geq \sum_{i=1}^{d} \lambda_i (yW)_i^2 = S_{jj}$$

(d)

The bound is tight, iff the eigenvector yW is of the form (100..00), i.e. if it corresponds to the first coordinate axis. If this is the case, then all the other eigenvectors also correspond to coordinate axes since W is orthogonal.