Machine Learning 1: Assignment 3

Tom Nick 340528
Niklas Gebauer 340942

Exercise 1

(a) In the following we want to minimize the objective function

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

subject to the constraint $\theta^T b = 0$, where $x_1, ..., x_k, \theta, b \in \mathbb{R}^d$.
Therefore we will define the Lagrangian function and set its gradient to zero in order to solve for λ and θ:

$$\mathcal{L}(\theta, \lambda) = J(\theta) + \lambda\theta^T b = \sum_{k=i}^{n} \|\theta - x_k\|^2 + \lambda\theta^T b$$
$$\nabla\mathcal{L} = \begin{pmatrix} [\sum_{k=i}^{n} 2(\theta - x_k)] + \lambda b \\ \theta^T b \end{pmatrix} \overset{!}{=} \vec{0}$$

We will at first solve the first entry of our gradient for θ:

$$[\sum_{k=i}^{n} 2(\theta - x_k)] + \lambda b = 0$$
$$\Leftrightarrow 2n\theta - 2[\sum_{k=i}^{n} x_k] + \lambda b = 0$$
$$\Leftrightarrow 2[\sum_{k=i}^{n} x_k] - \lambda b = 2n\theta$$
$$\Leftrightarrow \frac{1}{n}[\sum_{k=i}^{n} x_k] - \frac{1}{2n}\lambda b = \theta$$

Let's plug this into the second entry to solve for λ:

$$\theta^T b = 0$$
$$\Leftrightarrow (\frac{1}{n}[\sum_{k=i}^{n} x_k] - \frac{\lambda}{2n} b)^T b = 0$$
$$\Leftrightarrow (\frac{1}{n}[\sum_{k=i}^{n} x_k^T] - \frac{\lambda}{2n} b^T) b = 0$$
$$\Leftrightarrow \frac{1}{n}[\sum_{k=i}^{n} x_k^T] b - \frac{\lambda}{2n} b^T b = 0$$
$$\Leftrightarrow \frac{1}{n}[\sum_{k=i}^{n} x_k^T] b = \frac{\lambda}{2n} b^T b$$
$$\Leftrightarrow \frac{2}{b^T b}[\sum_{k=i}^{n} x_k^T] b = \lambda$$

Finally, we can plug this λ into our formula describing θ, yielding the result:

$$\begin{aligned}
\theta &= \frac{1}{n}[\sum_{k=i}^{n} x_k] - \frac{1}{2n}\lambda b \\
\Leftrightarrow \theta &= \frac{1}{n}[\sum_{k=i}^{n} x_k] - \frac{1}{2n}(\frac{2}{b^T b}[\sum_{k=i}^{n} x_k^T]b)b \\
\Leftrightarrow \theta &= \frac{1}{n}([\sum_{k=i}^{n} x_k] - \frac{[\sum_{k=i}^{n} x_k^T b]}{b^T b}b)
\end{aligned}$$

Similar to the minimal solution for θ of the unconstrained objective, we again find the the empirical mean in our solution. But this time we subtract a scaled version of the vector b. This means, that the solution still has to lie close to the empirical mean. If we imagine a plot with contour lines of $J(\theta)$ there has to be a 'valley' which is the empirical mean and a line cutting the surface which specifies the points where our constraint holds. The second term of our solution will move θ from total minimum (the empirical mean) to the point where the line lies deepest in the plane and where the lines and J's gradient both are perpendicular to the line. So we have the minimal solution that still fulfills the constraint.

(b) Now we will repeat the same procedure for the objective above with a different constraint ($\|\theta - c\|^2 = 1, c \in \mathbb{R}^d$):

$$\begin{aligned}
\mathcal{L}(\theta, \lambda) &= J(\theta) + \lambda\|\theta - c\|^2 - \lambda \\
\nabla\mathcal{L} &= \begin{pmatrix} [\sum_{k=i}^{n} 2(\theta - x_k)] + \lambda 2(\theta - c) \\ \|\theta - c\|^2 - 1 \end{pmatrix} \stackrel{!}{=} \vec{0}
\end{aligned}$$

Solve the first entry for θ:

$$\begin{aligned}
&[\sum_{k=i}^{n} 2(\theta - x_k)] + \lambda 2(\theta - c) = 0 \\
\Leftrightarrow\ & 2n\theta - 2[\sum_{k=i}^{n} x_k] + 2\lambda\theta - 2\lambda c = 0 \\
\Leftrightarrow\ & 2n\theta + 2\lambda\theta = 2[\sum_{k=i}^{n} x_k] + 2\lambda c \\
\Leftrightarrow\ & \theta = \frac{1}{n+\lambda}([\sum_{k=i}^{n} x_k] + \lambda c)
\end{aligned}$$

Let's plug this into the second entry to solve for λ:

$$
\begin{aligned}
\|\theta - c\|^2 - 1 = 0 \\
\Leftrightarrow \|\frac{1}{n+\lambda}([\sum_{k=i}^{n} x_k] + \lambda c) - c\|^2 - 1 = 0 \\
\Leftrightarrow \|\frac{1}{n+\lambda}[\sum_{k=i}^{n} x_k] + \frac{\lambda}{n+\lambda}c - \frac{n+\lambda}{n+\lambda}c\|^2 - 1 = 0 \\
\Leftrightarrow \|\frac{1}{n+\lambda}[\sum_{k=i}^{n} x_k] - \frac{n}{n+\lambda}c\|^2 = 1 \\
\Leftrightarrow \|\frac{1}{n+\lambda}([\sum_{k=i}^{n} x_k] - nc)\|^2 = 1 \\
\Leftrightarrow (\frac{1}{n+\lambda}([\sum_{k=i}^{n} x_k] - nc))^T(\frac{1}{n+\lambda}([\sum_{k=i}^{n} x_k] - nc)) = 1 \\
\Leftrightarrow \frac{1}{(n+\lambda)^2}([\sum_{k=i}^{n} x_k] - nc)^T([\sum_{k=i}^{n} x_k] - nc) = 1 \\
\Leftrightarrow \|[\sum_{k=i}^{n} x_k - c]\|^2 = (\lambda + n)^2 \\
\Leftrightarrow \|[\sum_{k=i}^{n} x_k - c]\|^2 = \lambda^2 + 2n\lambda + n^2 \\
\Leftrightarrow 0 = \lambda^2 + 2n\lambda + n^2 - \|[\sum_{k=i}^{n} x_k - c]\|^2
\end{aligned}
$$