Machine Learning I at TU Berlin

Assignment 3 - Group PTHGL

November 5, 2018

Exercise 1

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

(a)

The objective function $J(\theta)$ is given by

$$\begin{aligned} J(\theta) &= \sum_{k=1}^{n} \|\theta - x_k\|^2 \\ &= n\theta^T\theta - 2\theta\sum_{k=1}^{n} x_k + \theta\sum_{k=1}^{n} x_k x_k^T \quad . \end{aligned}$$

with the constraint $g(\theta) = \theta^T b = 0$, we get the Lagrange function

$$\mathcal{L}(\theta, \lambda) = n\theta^T\theta - 2\theta\sum_{k=1}^{n} x_k + \theta\sum_{k=1}^{n} x_k x_k^T - \lambda\theta^T b \quad .$$

By setting the partial derivative to zero we get the function for the

$$\frac{\partial\mathcal{L}(\theta,\lambda)}{\partial\theta} = 2n\theta - 2\underbrace{\sum_{k=1}^{n} x_k}_{n\overline{x}} - 2\lambda\theta^T b$$

$$\longrightarrow \theta = \frac{2n\overline{x} - \lambda b}{2n}$$

We then insert our result in the constraint for θ, and get a result for λ.

$$\begin{aligned} 0 &= \theta^T b \\ 0 &= b^T\lambda \\ 0 &= \frac{2nb^T\overline{x} - b^T\lambda b}{2n} \\ \longrightarrow \lambda &= \frac{2nb^T}{b^T b} \end{aligned}$$

Geometrically, the constrain $g(\theta)$ implies that θ is orthogonal to the vector b.

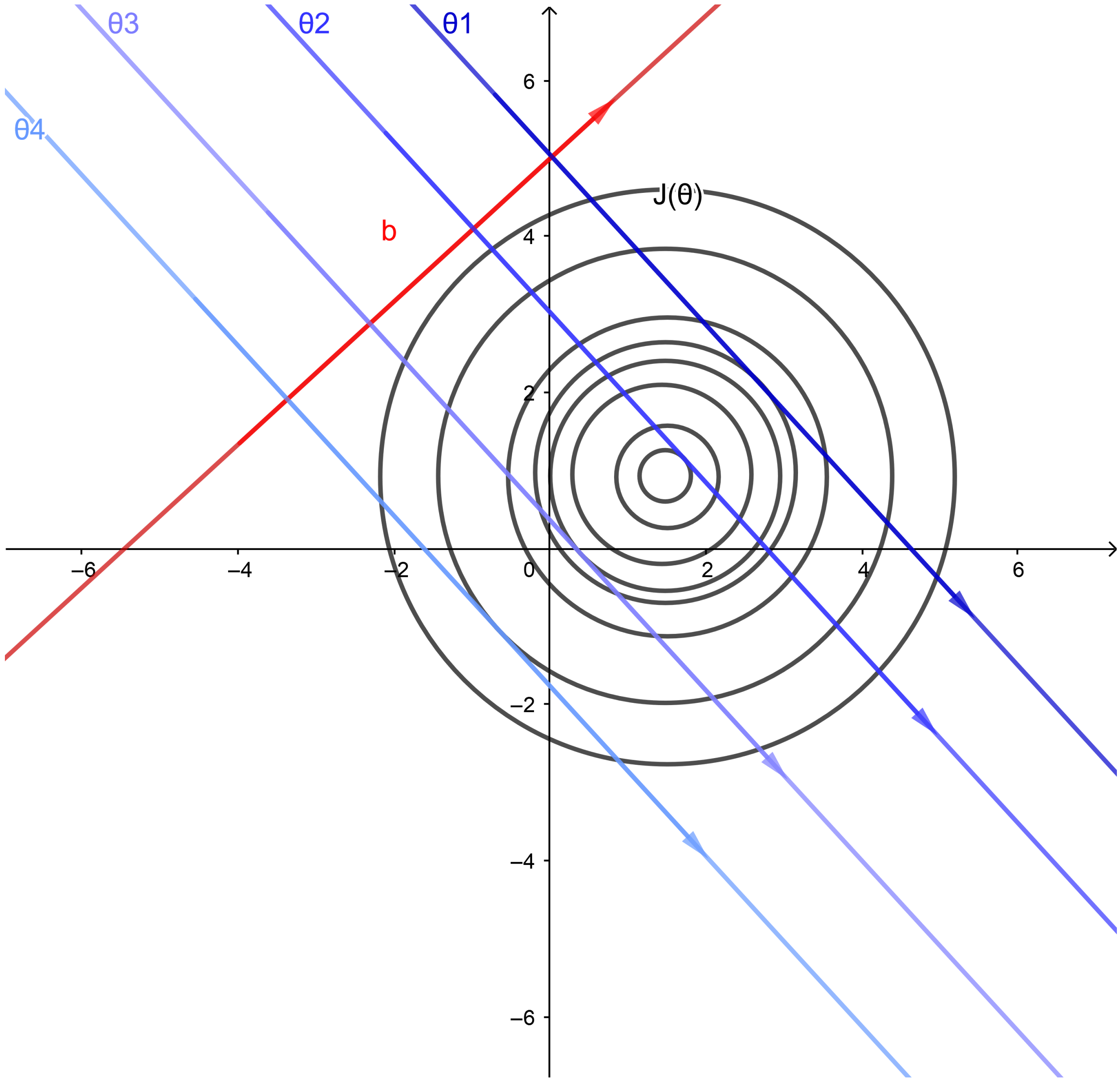

Figure 1: Geometrical interpretation of 1a. in a 2D plane.

(b)

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

with the constraint $g(\theta) = \|\theta - c\|^2 - 1 = 0$, we get the Lagrange function:

$$\mathcal{L}(\theta, \lambda) = \sum_{k=1}^{n} \|\theta - x_k\|^2 + \lambda(\|\theta - c\|^2 - 1).$$

Now we need to find the parameter θ that minimizes the Lagrange function $\frac{\partial \mathcal{L}(\theta,\lambda)}{\partial \theta} = 0$

$$\begin{aligned}
\frac{\partial \mathcal{L}(\theta,\lambda)}{\partial \theta} &= 2\sum_{k=1}^{n}(\theta - x_k) + 2\lambda(\theta - c) \\
&= 2n\theta - 2\underbrace{\sum_{k=1}^{n} x_k}_{n\overline{x}} + 2\lambda(\theta - c) = 2n(\theta - \overline{x}) + 2\lambda(\theta - c) \\
&= 2\theta(n+\lambda) - 2n\overline{x} - 2\lambda c = 0 \Longrightarrow \theta_{min} = \frac{n\overline{x} + \lambda c}{n+\lambda}
\end{aligned}$$

θ_{min} in $\|\theta - c\|^2 = 1$:

$$\begin{aligned}
\|\theta - c\|^2 &= \left\|\frac{n\overline{x} + \lambda c}{n+\lambda} - c\right\|^2 = \left\|\frac{n\overline{x} - cn}{n+\lambda}\right\| = 1 \qquad\qquad \|x\|^2 = x^T x \\
&= \left(\frac{n\overline{x} - cn}{n+\lambda}\right)^T \left(\frac{n\overline{x} - cn}{n+\lambda}\right) = \frac{(n\overline{x} - cn)^T(n\overline{x} - cn)}{(n+\lambda)^2} \\
&\longrightarrow n^2 \underbrace{(x^T x - c^T c)}_{\|x-c\|^2} = (n+\lambda)^2 \\
&\longrightarrow \lambda_{1/2} = -n \pm \sqrt{n^2 \|x - c\|^2}
\end{aligned}$$

The constrain $g(\theta) = \|\theta - c\|^2 - 1 = 0$ implies θ has a

Excercise 2

a)

The given scatter matrix $S = \sum_{k=1}^{n} (x_k - m)(x_k - m)^T$ has the following properties:

$$\begin{aligned}
(I)\ Tr(S) = \sum_{i=1}^{d} S_{ii} = \sum_{i=1} \lambda_i = \lambda_1 + \sum_{i=2} \lambda_i \\
\Rightarrow \lambda_1 = \sum_{i=1}^{d} S_{ii} - \sum_{i=2} \lambda_i
\end{aligned}$$

By proofing that $\sum_{i=2} \lambda_i \geq 0$ we can say that:

$$\begin{aligned}
\sum_{i=1}^{d} S_{ii} - \sum_{i=2} \lambda_i \leq \sum_{i=1}^{d} S_{ii} \\
\Leftrightarrow \lambda_1 \leq \sum_{i=1}^{d} S_{ii}
\end{aligned}$$

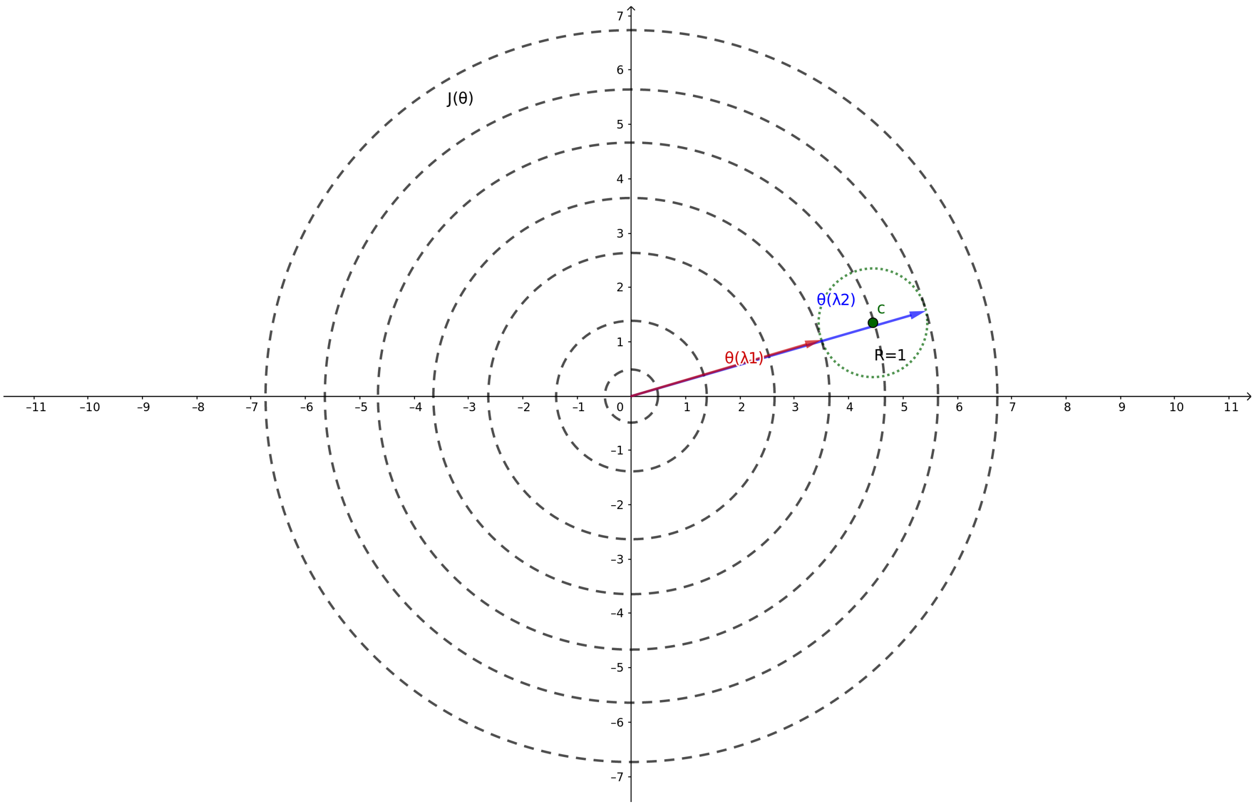

Figure 2: Geometrical interpretation of 1a. in a 2D plane .

Thus $\sum_{i=1}^{d} S_{ii}$ is an upper bound for λ_1.

Proof of $\sum_{i=2} \lambda_i \geq 0$:

A symmetric matrix S does have only positive or zero valued Eigenvalues λ iff it's positive semi-definite.

S is positive semi-definite iff: $z^T S z \geq 0$, for every non-zero column vector z of n real numbers.

$$z^T S z \Leftrightarrow \sum_{k=1}^{n} z^T (x_k - m)(x_k - m)^T z$$

$$\Leftrightarrow \sum_{k=1}^{n} z^T z (x_k - m)^T (x_k - m)$$

$$\Leftrightarrow \sum_{k=1}^{n} \|z\|^2 \|x_k - m\|^2 \overset{!}{\geq} 0$$

$\Rightarrow$ S is positive semi-definite

$\Rightarrow \sum_{i=1} \lambda_i \geq 0$

$\Rightarrow \sum_{i=1}^{d} S_{ii}$ is an upper bound for λ_1.

b)

The upper bound is tight, if the variance within the data set can be represented along one dimension.

That means that the variance of the dataset is represented by one single Eigenvalue with its corresponding Eigenvector λ_{σ^2} $\Rightarrow$ the value of all other Eigenvalues equals 0.

In that case equation (I) can be expressed as:

$$\sum_{i=1}^{d} S_{ii} = \sum_{i=1} \lambda_i = \lambda_1 + \sum_{i=2} \lambda_i$$
$$\Rightarrow \sum_{i=1}^{d} S_{ii} = \lambda_1$$

c)

From slides 3 page 26

d)

Exercise 3

$$w \longleftarrow \frac{Sw}{\|Sw\|}$$

(a)

The unconstrained objective function $J(\omega)$ is given by

$$J(\omega) = \|S\omega\| - \frac{1}{2}\omega^T S\omega$$

The gradient ascent

$$v \longleftarrow v + \gamma \frac{\partial(J)}{\partial(v)}$$

where v is given by

$$v = S^{\frac{1}{2}}\omega$$

$$
\begin{aligned}
\omega &= \frac{v}{S^{\frac{1}{2}}} \\
&and \\
\frac{\partial \omega}{\partial v} &= \frac{1}{S^{\frac{1}{2}}} \\
\frac{\partial J(\omega)}{\partial v} &= \frac{\partial J(\omega)}{\partial \omega} \cdot \frac{\partial \omega}{\partial v} \\
&= \big(\frac{SS\omega}{\|S\omega\|} - S\omega\big) \cdot S^{\frac{1}{2}} \\
&= \frac{S \cdot S^{\frac{1}{2}}\omega}{\|S\omega\|} - S^{\frac{1}{2}}\omega \\
&= \frac{Sv}{\|S\omega\|} - v \\
\Longrightarrow v &= v + \gamma\big(\frac{Sv}{\|S\omega\|} - v\big)
\end{aligned}
$$

if $\gamma = 1$

$$
\begin{aligned}
v \leftarrow v + \big(\frac{Sv}{\|S\omega\|} - v\big) \\
v \leftarrow \frac{Sv}{\|S\omega\|}
\end{aligned}
$$

(b)

The maximize of the objective $J(\omega)$ is a unit vector $\|\omega\| = 1$

$$
\begin{aligned}
\frac{\partial J(\omega)}{\partial \omega} &= 0 \\
\big(\frac{SS\omega}{\|S\omega\|} - S\omega\big) &= 0 \\
S\omega &= \omega\|S\omega\| \\
\|S\omega\| &= \|S\omega\| \cdot \|\omega\| \\
\longrightarrow \|\omega\| &= 1
\end{aligned}
$$

We cannot have to the maximize of the objective $J(\omega)$ more than one solution, so if $S\,' = \frac{S}{n}n$ then

$$
\begin{aligned}
\frac{S\omega}{\|S\omega\|} &= \frac{\frac{S}{n}n\omega}{\|\frac{S}{n}n\omega\|} \\
&= \frac{S\,'\omega\,'}{\|S\,'\omega\,'\|} \\
&\neq 1
\end{aligned}
$$