sheet03_theo_leo

November 5, 2018

1 Exercise sheet 03

2 Exercise 1: Lagrange Multipliers

Considering the function $J(\theta) = \sum_k^n ||\theta - x_k||^2$ with parameter $\theta \in \mathbb{R}^d$:
We define $\overline{x} = \frac{1}{n}\sum_k^n x_k$.

(a) Find θ that minimizes $J(\theta)$ under the constraint $\theta^T b = 0$ with $ b \in \mathbb{R}^d$:
$L(\theta, \lambda) = \sum_k^n ||\theta - x_k||^2 + \lambda \cdot (\theta^T b)$ and $\nabla L(\theta, \lambda) = \theta^T b + \sum_k^n \sum_i^d 2(\theta - x_k)_i + \lambda b$>
Setting $\nabla L(\theta, \lambda) = 0$ we obtain $\theta = \overline{x} + \frac{\lambda b}{2 \cdot n}$
Geometrical interpretation: The empirical mean with a shift directly proportional to b and inverse proportional to the number of samples n. Because of $\theta^T b = 0$, θ and b are perpendicular.

(b) Find θ that minimizes $J(\theta)$ under the constraint $||\theta - c||^2 = 1$ with $ c \in \mathbb{R}^d$:
$L(\theta, \lambda) = \sum_k^n ||\theta - x_k||^2 + \lambda \cdot (||\theta - c||^2 - 1)$ and $\nabla L(\theta, \lambda) = \sum_k^n \sum_i^d 2(\theta - x_k)_i + 2\lambda \sum_i^d (\theta - c)_i - 1 + \sum_i^d (\theta - c)_i^2$
Setting $\nabla L(\theta, \lambda) = 0$ we obtain $\theta = \frac{n}{n-1}\overline{x} - \frac{c}{n-1}$ with $||\theta - c||^2 = 1$
Geometrical interpretation: For large samples, θ approaches the empirical mean shifted by $\frac{c}{n-1}$. If also $c << n$, θ approaches the empirical mean. The distance between c and θ is hold constantly at 1.

3 Exercise 2: Bounds on Eigenvalues:

Dataset $x_n \in \mathbb{R}^d$, empirical mean $m = \frac{1}{n}\sum_k^n x_k$ and Scatter matrix $S = \sum_k^n (x_k - m)(x_k - m)^T$
With λ_1 beeing the largest eigenvalue of S and S_{ii} the diagonal elements of S

(a) Upper bounds of the eigenvalue λ_1:

With the trace beeing the sum over all eigenvalues the term $\sum_k^n S_{ii}$ is an upper bound for λ_1, because of the postive semidefiniteness of the scatter Matrix S.
proof: consideration of a random column vector a
$a^T S a = a^T (\sum_k^n (x_k - m)(x_k - m)^T) a = \sum_k^n (a^T (x_k - m))^2$
Since a^T, x_k and m are real valued, this is greater or equal to zero. Therefore all eigenvalues ar egreater or equal to zero.

(b) Dataconditions for which the upper bound is tight:

that would be that all the datavariation is along PCA1 and nothing around the Rest, eg. all other eigenvectors are zero.

(c) Lower bounds of the eigenvalue λ_1:

all positiv => minimum

(d) Dataconditions for which the lower bound is tight:

If there is no PCA(?)

4 Exercise 3: Iterative Principal Component Analysis

(a) Proof of the equality of the power iteration algorithm and the definition of the unconstrained objective $J(w) = ||Sw|| - 1/2w^T Sw$ and performing the gradient ascent $v \rightarrow v + \gamma \frac{\partial J}{\partial v}$ where $v = S^{1/2}$ assuming that γ is some learning rate and S is invertible.

\$ $\frac{\partial J}{\partial v = \frac{S \cdot w}{||S \cdot w||} - S\{1/2\} \cdot w \$ and therefore \$ v + \gamma(\frac{S \cdot w}{||S \cdot w||} - S\{1/2\} \cdot w) \rightarrow v \$ with v = S^{1/2} \cdot w}$
and for $\gamma \rightarrow 1$ we obtain $(\frac{S \cdot w}{||S \cdot w||} - S^{1/2} \cdot w + S^{1/2} \cdot w) \rightarrow v$ and $\frac{S \cdot w}{||S \cdot w||} \rightarrow v$.