Exercises for the course
Machine Learning 1
Winter semester 2018/19

Abteilung Maschinelles Lernen
Institut für Softwaretechnik und theoretische Informatik
Fakultät IV, Technische Universität Berlin
Prof. Dr. Klaus-Robert Müller
Email: klaus-robert.mueller@tu-berlin.de

Exercise Sheet 3

Exercise 1: Lagrange Multipliers (10+10 P)

Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in \mathbb{R}^d$ be a dataset of n samples. We consider the objective function

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

to be minimized with respect to the parameter $\boldsymbol{\theta} \in \mathbb{R}^d$. It can be shown that in absence of constraints for $\boldsymbol{\theta}$, the parameter $\boldsymbol{\theta}^*$ that minimizes this objective is given by the empirical mean $\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{k=1}^{n} \boldsymbol{x}_k$. However, this is not necessarily the case when the parameter $\boldsymbol{\theta}$ is constrained.

(a) Using the method of Lagrange multipliers, *find* the parameter $\boldsymbol{\theta}$ that minimizes $J(\boldsymbol{\theta})$ subject to the constraint $\boldsymbol{\theta}^\top \boldsymbol{b} = 0$, where $\boldsymbol{b} \in \mathbb{R}^d$. Give a geometrical interpretation to your solution.

(b) Using the same method, *find* the parameter $\boldsymbol{\theta}$ that minimizes $J(\boldsymbol{\theta})$ subject to $||\boldsymbol{\theta} - \boldsymbol{c}||^2 = 1$, where $\boldsymbol{c} \in \mathbb{R}^d$. Give a geometrical interpretation to your solution.

Exercise 2: Bounds on Eigenvalues (10+5+10+5 P)

We consider a dataset $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in \mathbb{R}^d$. The empirical mean $\boldsymbol{m}$, and the scatter matrix $\boldsymbol{S}$ are given by

$$\boldsymbol{m} = \frac{1}{n}\sum_{k=1}^{n} \boldsymbol{x}_k \qquad \text{and} \qquad \boldsymbol{S} = \sum_{k=1}^{n} (\boldsymbol{x}_k - \boldsymbol{m})(\boldsymbol{x}_k - \boldsymbol{m})^\top.$$

Let λ_1 be the largest eigenvalue of the matrix $\boldsymbol{S}$. The eigenvalue λ_1 quantifies the amount of variation in the data on the first principal component. Because computation of the full scatter matrix and respective eigenvalues can be slow, it can be useful to relate them to the diagonal elements of the scatter matrix $\{\boldsymbol{S}_{ii}\}$ that can be computed in linear time.

(a) *Show* that $\sum_{i=1}^{d} \boldsymbol{S}_{ii}$ is an upper bound to the eigenvalue λ_1.

(b) *State* the conditions on the data for which the upper bound is tight.

(c) *Show* that $\max_{i=1}^{d} \boldsymbol{S}_{ii}$ is a lower bound to the eigenvalue λ_1.

(d) *State* the conditions on the data for which the lower bound is tight.

Exercise 3: Iterative PCA (10+10 P)

When performing principal component analysis, computing the full eigendecomposition of the scatter matrix $\boldsymbol{S}$ is typically slow, and we are often only interested in the few first principal components. An efficient procedure to find the first eigenvector is the power iteration method, which starts with a random vector $\boldsymbol{w} \in \mathbb{R}^d$, and iteratively applies the parameter update

$$\boldsymbol{w} \leftarrow \frac{\boldsymbol{S}\boldsymbol{w}}{\|\boldsymbol{S}\boldsymbol{w}\|}$$

until some convergence criterion is met.

(a) *Show* that application of the power iteration method is equivalent to defining the unconstrained objective

$$J(\boldsymbol{w}) = \|\boldsymbol{S}\,\boldsymbol{w}\| - \frac{1}{2}\boldsymbol{w}^\top \boldsymbol{S}\boldsymbol{w}$$

and performing the gradient ascent $\boldsymbol{v} \leftarrow \boldsymbol{v} + \gamma\frac{\partial J}{\partial \boldsymbol{v}}$, where $\boldsymbol{v} = \boldsymbol{S}^{0.5}\boldsymbol{w}$ is a reparameterization of $\boldsymbol{v}$, for some learning rate γ. We assume that the matrix $\boldsymbol{S}$ is invertible.

(b) *Show* that a necessary condition for $\boldsymbol{w}$ to maximize the objective $J(\boldsymbol{w})$ is to be a unit vector (i.e. $\|\boldsymbol{w}\| = 1$).

Exercise 4: Programming (30 P)

Download the programming files on ISIS and follow the instructions.