Winter term 2024/25
U. von Luxburg
E. Günther/ K. Frohnapfel

# Assignment 04

## Mathematics for Machine Learning

Submission due Friday **15.11.2024, 23:59** via Ilias

Justify all your claims.

**Exercise 1** (**Principal Component Analysis**, 1 + 2 + 3 + 1 points)**.**
Consider the data matrix $X \in \mathbb{R}^{n \times d}$ representing $n$ data points (rows) and $d$ features (columns) and its covariance matrix $\Sigma = X^t X$. With principal component analysis (PCA) we want to find a data representation in a lower dimensional space $\tilde{X} \in \mathbb{R}^{n \times p}$ with $p < d$ features, that are not correlated but preserve as much of the variance as possible.

*Empirical variance and covariance:* For this exercise we use the empirical variance and covariance, which will be introduced later in the lecture. However, to solve this exercise you only need the following definition. Consider observations $a = (a_1, \dots, a_n)^t \in \mathbb{R}^n$ and $b = (b_1, \dots, b_n)^t \in \mathbb{R}^n$. If $a$ and $b$ are mean-centered, that is, $a_1 + ... + a_n = 0$ and $b_1 + ... + b_n = 0$, the empirical variance and covariance are defined by

$$\widehat{\mathrm{Var}}(a) = \frac{1}{n-1} \sum_{i=1}^{n} a_i^2$$
$$\widehat{\mathrm{Cov}}(a, b) = \frac{1}{n-1} \sum_{i=1}^{n} a_i b_i.$$

a) Prove that $\Sigma$ only has real eigenvalues.

b) Which vector $w^*$ solves the constrained maximization problem

$$\max_{\substack{w \in \mathbb{R}^d \\ \|w\|=1}} \widehat{\mathrm{Var}}(Xw)$$

and what is the maximal variance?
*Hint: You can assume that $Xw$ is a mean-centered vector.*

c) Which vector $v^*$ solves the constrained maximization problem

$$\max_{\substack{v \in \mathbb{R}^n \\ \|v\|=1}} \widehat{\mathrm{Var}}(Xv) \quad \text{subject to} \quad \widehat{\mathrm{Cov}}(Xv, Xw^*) = 0$$

and what is the maximal variance?.
*Hint: You can assume that $Xw^*$ and $Xv$ are mean-centered vectors.*

d) We call $Xw^*$ and $Xv^*$ the first and second principal components. How would you generalize this process to find the $k$-th principal component for $k \leq p < d$ and how do you find $\tilde{X}$? What would happen if you chose $p = d$? (no proof necessary)

**Exercise 2** (**Singular value decomposition**, 2+1+2+2 points)**.**

a) Compute the singular values of $M = \begin{pmatrix} -4 & 2 \\ 1 & 2 \\ 3 & 2 \end{pmatrix} \in \mathbb{R}^{3\times 2}$.

b) For a matrix $A \in \mathbb{R}^{m\times n}$, prove that $A$ and $A^t$ have the same singular values.

c) Prove that the Frobenius norm defined by $\|A\|_F = \sqrt{\operatorname{tr}(A^t A)}$ can be computed as $\|A\|_F = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2}$, where $\sigma_i$ are the singular values of $A \in \mathbb{R}^{m\times n}$.

d) Consider a matrix $A \in \mathbb{R}^{m\times n}$ with singular value decomposition $A = U\Sigma V^t$, where $U \in \mathbb{R}^{m\times m}, V \in \mathbb{R}^{n\times n}$ are orthonormal and $\Sigma \in \mathbb{R}^{m\times n}$ is diagonal. The matrix $\Sigma^\# \in \mathbb{R}^{n\times m}$ is obtained from $\Sigma$ by transposing and inverting every non-zero element, as we saw in the lecture. Prove that the matrix $A^\# = V\Sigma^\# U^t$ is indeed a pseudo-inverse of $A$. That is, it satisfies all the properties of a pseudo-inverse.

**Exercise 3** (**Condition number**, 3+3 points)**.** For an invertible matrix $A \in \mathbb{R}^{n\times n}$ we define the *condition number*

$$\kappa(A) := \|A\|_2 \cdot \left\|A^{-1}\right\|_2 .$$

It is used to measure how stable the solution of a linear system of equations is under the addition of noise. Consider the linear system of equations $Ax = b$ for some $b \in \mathbb{R}^n$.

a) Consider some noise $\Delta b \in \mathbb{R}^n$ on $b$ and the solution $x + \Delta x$ to the noisy equation, i.e. $A(x + \Delta x) = b + \Delta b$. Prove that the relative error can be bounded by

$$\frac{||\Delta x||}{||x||} \leq \kappa(A) \cdot \frac{\|\Delta b\|}{\|b\|},$$

where $\|\cdot\|$ denotes the Euclidean norm.
*Hint: Prove and use the inequality $\|Av\| \leq \|A\|_2 \cdot \|v\|$. for an arbitrary vector $v \in \mathbb{R}^n$.*

b) Find examples for $A$, $b$ and noise $\Delta b$ such that $\|\Delta b\|/\|b\| \leq 0.001$ and $\|\Delta x\|/\|x\| \geq 1$.