Winter term 2026/27
U. von Luxburg
E. Günther/ K. Frohnapfel

Assignment 4

Mathematics for Machine Learning

Prepare for Tutorial of **Week 6 (16.-20.11.2026)**

Justify all your claims.

Exercise 1 (Spectral clustering).
Let $W \in \mathbb{R}^{n\times n}$ be a symmetric matrix with non-negative entries and $D \in \mathbb{R}^{n\times n}$ the diagonal matrix which contains the row sums of W, that is, $d_{i,i} = \sum_{j=1}^{n} w_{i,j}$. Define $L := D - W$.

a) Prove that for all $x \in \mathbb{R}^n$ it holds

$$x^t L x = \frac{1}{2} \sum_{i,j=1}^{n} w_{i,j}(x_i - x_j)^2 .$$

b) Conclude that L is symmetric and positive semi-definite.

c) Show that the vector of constant ones $\mathbb{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n$ is an eigenvector of L.

d) Solve the constrained minimization problem

$$\min_{\substack{x\in\mathbb{R}^n \\ \|x\|=1}} x^t L x \quad \text{subject to } \langle x, \mathbb{1}\rangle = 0 .$$

Note: This exercise introduces the basic principle behind spectral clustering. The matrix W is called the adjacency matrix and holds the edge weights of a graph as entries. D holds the node degrees on its diagonal and is therefore called the degree matrix and L is called the (unnormalized) graph Laplacian matrix. The minimization problem in d) is a relaxation of the min cut problem, which tries to find the minimal number of edges that you have to remove to separate a graph into two subgraphs.

Exercise 2 (Principal Component Analysis).
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 $a = (a_1, \ldots, a_n)^t \in \mathbb{R}^n$ and $b = (b_1, \ldots, b_n)^t \in \mathbb{R}^n$. 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

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

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

© 2026 University of Tübingen

a) Prove that Σ 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

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

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

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

where $\|\cdot\|_2$ is the spectral norm from the lecture. 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 $\tilde{x}$ to the noisy equation, i.e. $A\tilde{x} = b + \Delta b$. Let $\Delta x := \tilde{x} - x$ be the absolute error. 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 Δb such that $\|\Delta b\|/\|b\| \leq 0.001$ and $\|\Delta x\|/\|x\| \geq 1$.