Exercises for the course
Machine Learning for Data Science
Winter Semester 2024/25

G. Montavon
Institute of Computer Science
Department of Mathematics and Computer Science
Freie Universität Berlin

Exercise Sheet 3 (theory part)

Exercise 1: Principal Component Analysis (15 + 15 P)

We consider a dataset $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N \in \mathbb{R}^d$. Principal component analysis searches for a unit vector $\boldsymbol{u} \in \mathbb{R}^d$ such that projecting the data on that vector produces a distribution with maximum variance. Such vector can be found by solving the optimization problem:

$$\arg\max_{\boldsymbol{u}} \frac{1}{N}\sum_{i=1}^{N}\left[\boldsymbol{u}^\top(\boldsymbol{x}_i - \boldsymbol{m})\right]^2 \quad \text{with} \quad \|\boldsymbol{u}\|^2 = 1 \quad \text{and} \quad \boldsymbol{m} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i.$$

(a) *Show* that the problem above can be rewritten as

$$\arg\max_{\boldsymbol{u}} \ \boldsymbol{u}^\top \boldsymbol{\Sigma} \boldsymbol{u} \quad \text{with} \quad \|\boldsymbol{u}\|^2 = 1$$

where $\boldsymbol{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i - \boldsymbol{m})(\boldsymbol{x}_i - \boldsymbol{m})^\top$ is the covariance matrix.

(b) *Show* using the method of Lagrange multipliers that the problem above can be reformulated as solving the eigenvalue problem

$$\boldsymbol{\Sigma}\boldsymbol{u} = \lambda \boldsymbol{u}$$

and retaining the eigenvector $\boldsymbol{u}$ associated to the highest eigenvalue λ.

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

Let λ_1 denote the largest eigenvalue of the matrix $\boldsymbol{\Sigma}$. The eigenvalue λ_1 measures the variance of the data when projected on the first principal component. We study how the latter can be bounded with the diagonal elements of the matrix $\boldsymbol{\Sigma}$.

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

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

Exercise 3: Iterative PCA (5 + 10 + 5 P)

When performing principal component analysis, computing the full eigendecomposition of the covariance matrix $\boldsymbol{\Sigma}$ is typically slow, and we are often only interested in the first principal components. An efficient procedure to find the first principal component is *power iteration.* It starts with a random unit vector $\boldsymbol{w}^{(0)} \in \mathbb{R}^d$, and iteratively applies the parameter update

$$\boldsymbol{w}^{(t+1)} = \boldsymbol{\Sigma}\boldsymbol{w}^{(t)} \big/ \|\boldsymbol{\Sigma}\boldsymbol{w}^{(t)}\|$$

until some convergence criterion is met. Here, we would like to show the exponential convergence of power iteration. For this, we look at the error terms

$$\mathcal{E}_k(\boldsymbol{w}) = \left|\frac{\boldsymbol{w}^\top \boldsymbol{u}_k}{\boldsymbol{w}^\top \boldsymbol{u}_1}\right| \qquad \text{with} \quad k = 2, \dots, d,$$

will demonstrate that they all converge to zero as $\boldsymbol{w}$ approaches the eigenvector $\boldsymbol{u}_1$ and becomes orthogonal to other eigenvectors. To demonstrate this, we proceed in three steps:

(a) Show that $\mathcal{E}_k(\boldsymbol{w}^{(t+1)}) = \left|\dfrac{{\boldsymbol{w}^{(t)}}^\top \boldsymbol{\Sigma}\boldsymbol{u}_k}{{\boldsymbol{w}^{(t)}}^\top \boldsymbol{\Sigma}\boldsymbol{u}_1}\right|$.

(b) Starting from the result above, show that $\mathcal{E}_k(\boldsymbol{w}^{(t+1)}) = |\lambda_k/\lambda_1| \cdot \mathcal{E}_k(\boldsymbol{w}^{(t)})$ *(Hint: to show this, it is useful to recall that the covariance is linked to its eigenvectors and eigenvalues through the equation* $\boldsymbol{\Sigma}\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i$*.)*

(c) Starting from the result above, show that $\mathcal{E}_k(\boldsymbol{w}^{(T)}) = |\lambda_k/\lambda_1|^T \cdot \mathcal{E}_k(\boldsymbol{w}^{(0)})$, i.e. the convergence of the algorithm is exponential with the number of time steps T.