Universität des Saarlandes
Prof. Dr. Dietrich Klakow
Lehrstuhl für Signalverarbeitung
NNTI Winter Term 2024/2025

Exercise Sheet 2

Linear Algebra

Deadline: 6.11.2024 23:59

Guidelines: You are expected to work in a group of 2-3 students. While submitting the assignments, please make sure to include the following information for all our teammates in your PDF/python script:

Name:

Student ID (matriculation number):

Email:

Your submissions should be zipped as **Name1_id1_Name2_id2_Name3_id3.zip** when you have multiple files. For assignments where you are submitting a single file, use the **same naming convention** without creating a zip. For any clarification, please reach out to us on the **CMS Forum**.
Note that the above instructions are mandatory. If you are not following them, tutors can decide not to correct your exercise.

Exercise 2.1 - Linear Algebra (1+1+0.5 points)

a) Use that every matrix $A \in \mathbb{R}^{m \times n}$ can be decomposed into $A = UDV^T$, where V are the eigenvectors of $A^T A$, to show that the singular values of A (i.e. the values along the diagonal of D) are the square roots of the eigenvalues of $A^T A$.

b) Compute the eigenvalues and eigenvectors of the matrix $A = \begin{bmatrix} 9 & 0 & 0 \\ 5 & 2 & 0 \\ 7 & 4 & 8 \end{bmatrix}$

c) Given an invertible real symmetric matrix B, what is the relationship between the eigenvalues and eigenvectors of B and B^{-1}? What about B and B^T?

Exercise 2.2 - Datasets and Dataloaders (4 points)

This exercise will help you get familiar with PyTorch Dataset and Dataloader classes for the next assignments. The goal of the exercise is to implement your own Dataset and Dataloader classes for the datasets Dataset1 and Dataset2.

Dataset1 is a dataset of tokenized text data from sms_spam. The features are the sequence of tokenized words and the label is 0 or 1 for "not spam" and "spam". Be wary that you might need to pad the token sequences to make the number of features consistent.

Dataset2 is a csv file of tabular data (the number of columns is fixed) comprising symptoms and prognosis of 11 vector-borne deseases. The symptoms are the features, while the

prognosis is the label. You are required to perform one-hot encoding of the label, to transform it into numerical values. For more information on one-hot encoding, check out one-hot encoding.

Please make sure you do the following for both:

- You will be required to split your data into 2 datasets, one for the training and one for the testing
- Your custom dataset class should subclass the Dataset class from PyTorch
- Your dataset class should have the required function(s) to iterate over the requested dataset i.e. len, iter, init etc.
- Your custom dataloader class should subclass the Dataloader class from PyTorch
- Your implementation of the dataloader class should work with the objects returned by your custom dataloader class
- Your dataloader should return a tuple of input/output

You are encouraged to lookup the source code of Dataloader and Dataset class in PyTorch. The main aim of this exercise is to help you understand the working of Dataset and Dataloader classes as we shall be using them in future assignments.

Exercise 2.3 - Eigencats (3.5 points)

See the accompanying jupyter notebook.