UNIVERSITÄT DES SAARLANDES
Prof. Dr. Dietrich Klakow
Lehrstuhl für Signalverarbeitung
NNTI Winter Term 2024/2025

Exercise Sheet 3

Machine Learning Basics

Deadline: 13.11.2024 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**. These instructions are mandatory. If you are not following them, tutors can decide not to correct your exercise.
Please note:

- **Notational clarification:** In this assignment, x,y,z denote scalar values; **x**, **y**, **z** denote vectors; and X, Y, Z denote matrices.
- Ex 3.1-3.3 are written assignments, please submit a pdf (written using Latex) with the **names, matriculation IDs and emails** of all team members for this part. In case you are not familiar with Latex, clearly written handwritten submissions are also accepted, but we strongly encourage pdfs written using Latex.
- Ex 3.4 and 3.5 are programming assignments, you can write your code in the supplied notebooks and submit them. Don't forget to put in your **names, matriculation IDs and emails**.
- Submit the pdfs and notebooks together in a zip file in CMS. No need to submit any datasets.

Exercise 3.1 - Linear Regression ((1+0.25+0.25)+0.5+0.5 points)

Linear regression aims to model the relationship between the matrix of data points X and the label vector $\mathbf{y}$. Considering a dataset of N data points, with each having n features, this relationship is formulated using a linear equation of the form:

$$y^{(i)} = \mathbf{w}^T\mathbf{x}^{(i)} = b + w_1 x_1^{(i)} + \cdots + w_n x_n^{(i)}; \quad (\text{ for a single data point})$$

or,

$$\mathbf{y} = X\mathbf{w}; \quad (\text{in matrix notation for the entire dataset})$$

where $\mathbf{w}$ represents the vector of regression parameters. Here, $X_{N\times(n+1)} = \begin{bmatrix} 1 & x_1^{(1)} & \cdots & x_n^{(1)} \\ 1 & x_1^{(2)} & \cdots & x_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(N)} & \cdots & x_n^{(N)} \end{bmatrix} =$

$\begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix}$, $\mathbf{y}_{N\times 1} = \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(N)} \end{bmatrix}$, and $\mathbf{w}_{(n+1)\times 1} = \begin{bmatrix} b \\ w_1 \\ \vdots \\ w_n \end{bmatrix}$.

Here, we include the bias term b as part of $\mathbf{w}$, so it has dimension $(n+1)$. The bias is also represented by the coefficient 1 in each data vector $\mathbf{x}_i$.

The optimal parameters that satisfy this equation given the data minimize the cost function:

$$f(\mathbf{w}) = \frac{1}{N}\|X\mathbf{w} - \mathbf{y}\|_2^2$$

a) This least-squares formulation has a closed-form analytical solution called the normal equation. Derive the formulation of the normal equation, clearly stating any lemmas or assumptions used in your derivation.

   Now, answer the following questions:

   i. What are the problems with using this normal equation method?

   ii. Does this work when the data matrix X is not invertible? If yes, how so?

b) Consider the following dataset $\mathcal{D}_1$:

x	9.0	2.0	6.0	1.0	8.0
y	1.0	0.0	3.0	0.0	1.0

   Using the normal equation derived above, find the optimal parameters $\mathbf{w}$ given the dataset $\mathcal{D}_1$.

c) Let us assume we already have the parameters $\mathbf{w}_{\mathcal{D}_1}^T = \begin{bmatrix} b_1 & w_1 \end{bmatrix}$ that minimize the mean-squared error (MSE) for $\mathcal{D}_1$. Now, consider another dataset $\mathcal{D}_2$:

x	9.0+γ	2.0+γ	6.0+γ	1.0+γ	8.0+γ
y	1.0+η	0.0+η	3.0+η	0.0+η	1.0+η

   where $\gamma, \eta > 0$ and $w_1\gamma \neq \eta$. Let $\mathbf{w}_{\mathcal{D}_2}^T = \begin{bmatrix} b_2 & w_2 \end{bmatrix}$ be the parameters that minimize the MSE for $\mathcal{D}_2$. Which of the cases listed below hold in this case? Explain your reasoning.

   a) $w_1 = w_2, b_1 = b_2$

   b) $w_1 \neq w_2, b_1 = b_2$

   c) $w_1 = w_2, b_1 \neq b_2$

   d) $w_1 \neq w_2, b_1 \neq b_2$

Exercise 3.2 - Principal Component Analysis (PCA)-I (0.5 + 0.5 points)

Please answer the following questions in two to three sentences:

a) Why is normalization an important step in PCA?

b) Give an example of an instance when PCA performs badly.

Exercise 3.3 - Principal Component Analysis (PCA)-II (0.5 + 0.5 + 0.5 points)

Let's assume we've performed PCA on the toy dataset shown in Table 1:

Table 1: Dataset

Row	**X1**	**X2**	**X3**	**X4**
1	0.49	0.07	0.12	-1.19
2	-0.35	1.14	0.18	0.57
3	-0.44	0.29	-0.85	0.30
4	0.65	-0.42	-0.30	-0.22
5	1.15	-0.44	0.77	0.98
6	0.45	0.14	-0.02	0.86

And we've obtained the principal components as shown in Table 2:

Table 2: Principal Components of the dataset

PC1	**PC2**	**PC3**	**PC4**
0.69	-0.24	0.03	-0.7
-0.49	0.33	0.58	-0.55
-0.45	0.06	0.76	0.45
0.32	-0.9	-0.26	-0.02

Which correspond to the following eigenvalues:

$$[0.739, 0.685, 0.239, 0.004]$$

Answer the following questions:

a) Why are there only 4 principal components?

b) How much of the variance in the data is preserved by the first two principal components?

c) How much of the variance in the data is preserved by the first and third principal components together?

Exercise 3.4 - Image denoising using PCA (5 points)

See attached notebook

Exercise 3.5 - Polynomial Regression (Bonus) (3 points)

See attached notebook