Universität des Saarlandes
Prof. Dr. Dietrich Klakow
Lehrstuhl für Signalverarbeitung
NNTI Winter Term 2023/2024

Exercise Sheet 3

Machine Learning Basics

Deadline: 23.11.2023 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 3.1 - Linear Regression (0.25+0.25+1+0.5 points)

Linear regression is about finding the line that best fits a collection of data. Where X represents the features, w the weights, b the noise and Y being the intercept, giving us an equation of the form:

$$Y = wX + b$$

a) Calculate w and b for the line that passes through the points (2, 4) and (4, 5).

b) Does the point (3, 6) lie on the line from **a)** above? Demonstrate your answer analytically.

c) Given the Mean squared error equation:

$$\text{MSE}_{\text{train}} = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}^{(i)}_{\text{train}} - y^{(i)}_{\text{train}})^2 \tag{1}$$

Minimize the MSE by finding the value of w where the gradient is equal to zero. Highlight every step of your proof and clearly state the formula and reasoning in each step.

d) The equation for Mean Squared Error only uses the difference between the y values. Why does it not take into account distances in the x direction instead?

Exercise 3.2 - PCA as Autoencoder (1 + 3 points)

As you have seen in the lecture, we can view PCA as a linear autoencoder. In this exercise, you will explore this connection in more detail. In particular, you will prove that the principal components minimize the reconstruction error 2. You can find more details in chapter 2.12 of [1].

a) Show that for a single sample $\boldsymbol{x}$ the optimal coding vector $\boldsymbol{c}$ that minimizes the reconstruction error of a linear decoder $\mathbf{D}$ with orthogonal columns, is given by $\mathbf{D}^T\boldsymbol{x}$. In other words, show that $\boldsymbol{c} = \mathbf{D}^T\boldsymbol{x}$ solves

$$\underset{\boldsymbol{c}}{\operatorname{argmin}} \, ||\boldsymbol{x} - \mathbf{D}\boldsymbol{c}||_2^2, \text{ subject to } \mathbf{D}^T\mathbf{D} = \mathbf{I}_l,$$

where $\mathbf{I}_l$ is a $m \times m$ matrix with ones along the first l diagonal entries and zeros everywhere else.

b) Argue by proof of induction that we can minimize the reconstruction error 2 by choosing the columns of $\mathbf{D}$ to be the eigenvectors of $\mathbf{X}^T\mathbf{X}$ corresponding to the largest eigenvalues, i.e. the principal components. To do so, first show that

$$\underset{\mathbf{D}}{\operatorname{argmin}} ||\mathbf{X} - \mathbf{X}\mathbf{D}\mathbf{D}^T||_F^2 = \underset{\mathbf{D}}{\operatorname{argmax}} \sum_{i=1}^{m} \mathbf{D}_{\cdot i}^T\mathbf{X}^T\mathbf{X}\mathbf{D}_{\cdot i},$$

where $\mathbf{D}_{\cdot i}$ is the ith column of $\mathbf{D}$. Using this, show that the eigenvector corresponding to the largest eigenvalue, i.e. the first principal component solves

$$\underset{d}{\operatorname{argmin}} \, ||\mathbf{X} - \mathbf{X}\boldsymbol{d}\boldsymbol{d}^T||_F \quad \text{subject to } \boldsymbol{d}^T\boldsymbol{d} = 1.$$

Afterwards, using this as a base case, proof by induction that the eigenvector corresponding to the nth largest eigenvalue, i.e. the nth principal component solves

$$\underset{d}{\operatorname{argmin}} \, ||\mathbf{X} - \mathbf{X}\boldsymbol{d}\boldsymbol{d}^T||_F \quad \text{subject to } \boldsymbol{d}^T\boldsymbol{d} = 1, \boldsymbol{d}^T\boldsymbol{d}_i = 0 \text{ for } i = 1, \ldots, n-1,$$

where $\boldsymbol{d}_i$ is the ith principal component. In other words, prove that the solution to

$$\underset{D}{\operatorname{argmin}} \, ||\mathbf{X} - \mathbf{X}\mathbf{D}\mathbf{D}^T||_F \quad \text{subject to } \mathbf{D}^T\mathbf{D} = \mathbf{I}_l, \tag{2}$$

is the matrix $\mathbf{D}$ whose columns are the eigenvectors of $\mathbf{X}^T\mathbf{X}$.

Hint: The trace operator is defined as $\operatorname{Tr}(\mathbf{A}) = \sum_{i=1}^{m} \mathbf{A}_{ii}$ and fulfills $\operatorname{Tr}(\mathbf{AB}) = \operatorname{Tr}(\mathbf{BA})$ for arbitrary matrices $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{B} \in \mathbb{R}^{n\times m}$.

Hint: $||\mathbf{A}||_F$ is the frobenius norm which fulfills $||\mathbf{A}||_F^2 = \operatorname{Tr}(\mathbf{A}^T\mathbf{A})$

Exercise 3.3 - PCA (4 points)

See the accompanying jupyter notebook.

References

[1] Ian Goodfellow and Yoshua Bengio and Aaron Courville. 2016. Deep Learning. MIT Press.