Universität des Saarlandes
Prof. Dr. Dietrich Klakow
Lehrstuhl für Signalverarbeitung
NNTI Winter Term 2022/2023

Exercise Sheet 3

Machine Learning Basics

Deadline: 30.11.2022 08:00

Exercise 3.1 - Eigen Value Decomposition (0.25 + 0.25 + 0.5 points)

Consider the following matrix:

$$M = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

a) Is the matrix M symmetric? What does it imply for its eigendecomposition?

b) Is the matrix M Singular? What does it imply for its eigendecomposition?

c) Find the eigendecomposition of M?

Exercise 3.2 - Linear Regression (0.5 + 0.5 + 1 + 2 points)

a) Briefly discuss the differences between Accuracy, Precision, Recall, F1. Give examples of situations when you would prefer one over the others. (max 100 words)

b) Suppose we have a dataset that contains the financial status of startups that have recently entered the stock market with three regressors, X_1 = "raised funds" (in millions of dollars), X_2 = "initial stock value", X_3 = "debt" (in millions of dollars). The variable of interest Y is the company value after a year (in millions of dollars). Suppose we use least squares to fit the model :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 ln X_{i2} + \beta_3 X_{i3}$$

where $\beta_0 = 10, \beta_1 = 10, \beta_2 = 0.5, \beta_3 = -5$.

Elaborate on the correctness of these statements :

a) One unit change X_1 causes a 1000 percent change in Y.

b) One unit change in X_2 causes a 50 percent change in Y.

c) 100 percent change in X_2 causes a 50 percent change in Y.

d) Is there any meaningful interpretation of the bias term β_0.

c) *The aim of this exercise is to deepen your understanding of regression and relate it to other statistical measures such as covariance and correlation.*
Consider two data series, $X = (x_1, x_2, ..., x_n)$ and $Y = (y_1, y_2, ..., y_n)$, with $\bar{X} = 0$ and $\bar{Y}$ = 0. We use linear regression (ordinary least squares) to regress Y against X (without fitting any intercept), as in Y = aX + ϵ where ϵ denotes a series of error terms.
Derive the value of the regression coefficient a in terms of the standard deviations σ_X and σ_Y and the correlation ρ_{XY} between the two data series.

d) *In this exercise we will show that adding noise to a data point is the same as adding regularization to your mean squared error cost function.*
Given a training dataset $D = \{x_i, y_i\}_n^{i=1}$ where $x_i \in R^d, y_i \in R$ and the mean squared error cost function $f(x, y; w) = \frac{1}{n}\Sigma_{i=1}^n (y_i - \langle w, x_i \rangle)^2$.
Consider adding a noise term ϵ_i where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ to each data sample in the training set x_i such that $\mathbb{E}[\sigma_i] = 0$ and $\mathbb{E}[\sigma_i \sigma_j] = \sigma^2$. We will assume $n \to \infty$. Thus, we write our cost function $f(x, y; w) = \frac{1}{n}\Sigma_{i=1}^n (y_i - \langle w, x_i \rangle)^2$ as $f(x, y; w) = \mathbb{E}[(y_i - \langle w, x_i \rangle)^2]$.
Prove that the following equation holds:

$$\mathbb{E}[(y_i \langle w, x_i + \epsilon_i \rangle)^2] = \mathbb{E}[(y_i - \langle w, x_i \rangle)^2] + \sigma^2 \sum_{i=1}^{d} w_i^2$$

Exercise 3.3 - Model Capacity, Overfitting and Underfitting (0.5 + 1 + 2.5 points)

a) Since we want good generalization (test) performance on data even though we learn only with available (train) data therefore, model capacity must be appropriate to the true complexity of the task. Higher capacity for a simpler task can lead to overfitting and low capacity for a complex task can lead to underfitting. Therefore when looking for an optimal capacity we generally make a Bias and Variance trade-off. More conclusively, the expected test mean square error (MSE), for a given value x, can always be decomposed into the sum of three fundamental quantities: the variance of $\hat{f}(x)$, the squared bias of $\hat{f}(x)$ and the variance of the error terms ε.
(Reading ISLR: section 2.2.2 The Bias-Variance Trade-off)

$$\mathbb{E}[(y - \hat{f}(x))]^2 = Var(\hat{f}(x)) + [Bias(\hat{f}(x))]^2 + Var(\varepsilon).$$

Give explanation of the terms Bias and Variance and how does they relate to Overfitting and Underfitting? (max 150 words)

b) Ridge Regression:
In the slide "Fitting with Regularization" a regularization term is introduced to the least squares loss to avoid over/underfitting. This is also known as ridge regression.

$$J(w) = MSE_{train} + \lambda w^T w$$

Here w are the weights and λ is the regularization parameter. Provide a closed form solution for w which minimizes this loss. Please show all steps in the solution.

c) See the attached .ipynb file

Exercise 3.4 - Validation set and Cross Validation (0.25 + 0.5 + 0.25 points)

We split our dataset and allocated a portion of it for validation purposes. Such a process of setting aside a portion of the dataset is known as holdout method and the dataset is called holdout dataset.

a) Why would one need cross-validation instead of the holdout method? Explain two scenarios where cross-validation is needed. (max 100 words)

b) Assume that we want to find an optimal capacity of our model for the task of linear regression, with possible choices for the order of the polynomial as: 1, 5, 9.
Now, assume that we want to do this hyperparameter selection using k-fold cross-validation (with $k = 5$), instead of the *holdout method*. Given this setting, explain the steps that are involved in this 5-fold cross-validation, along with a brief explanation. Also, explain how would you compute a single final score (e.g. MSE) for each of the hyperparameters (i.e. order of the polynomial), so that we can compare the performance of these models and choose the best one. (max 150 words)

c) You are given a dataset D = (x_i, y_i) with i=1..n . Assume that you ran leave-one-out cross-validation on this dataset and calculated the cross-validation mean squared error (MSE) and found that it is equal to 35. Now, you randomly shuffle the dataset D and re-run leave-one-out cross-validation and calculate the new MSE. How is the new MSE related to the previous MSE which was 35? Clearly justify your reasoning. (max 50 words)

Submission instructions

The following instructions are mandatory. If you are not following them, tutors can decide to not correct your exercise.

- Please submit the assignment as a **team of two to three** students.
- Write the Microsoft Teams user name, student id and the name of each member of your team on your submission.
- Hand in zip file containing a `single` PDF with your solutions and the completed ipython notebook. Do not include any data or cache files (e.g. __pycache__).
- Important: please name the submitted zip folder and files inside using the format: **Name1_id1_Name2_id_Name3_id3**.
- Your assignment solution must be uploaded by only **one** of your team members to the "Assignments" tab of the tutorial team (in **Microsoft Teams**). Please remember to press the **Hand In** button after uploading your work.
- If you have any trouble with the submission, contact your tutor **before** the deadline.