W3WI DS304 Applied Machine Learning Fundamentals

Exercise Sheet # 3 - Linear Regression

Question 1 (Matrix operations)

Let $\boldsymbol{X} \in \mathbb{R}^{n \times (m+1)}$, $\boldsymbol{y} \in \mathbb{R}^{n}$, and $\boldsymbol{\theta} \in \mathbb{R}^{m+1}$ be given. What is the result of the following operations (matrix, vector, or scalar), and what is the dimension?

- $\boldsymbol{X}^{\intercal}\boldsymbol{X}$
- $\boldsymbol{X}^{\intercal}\boldsymbol{y}$
- $\boldsymbol{y}^{\intercal}\boldsymbol{y}$
- $\|\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y}\|^2$

Question 2 EX 2022 (Normal equation)

Let the training dataset

$$\mathcal{D}_{\text{train}} = \{(1,2),(2,1),(3,3)\}$$

be given. Each training example is a tuple of the form (x, y), where $x \in \mathbb{R}$ is the only feature, and $y \in \mathbb{R}$ the corresponding label. Please work through the following questions:

1. Compute the optimal model parameters $\boldsymbol{\theta}^{\star}$ using the **normal equation**

$$\boldsymbol{\theta}^{\star} = \left(\boldsymbol{X}^{\intercal}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\intercal}\boldsymbol{y}.$$

   Do not apply any regularization. **Hint:** Let $\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2\times 2}$ be an invertible matrix. Its inverse is given by $\boldsymbol{A}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

2. Figure 1 below plots the training dataset $\mathcal{D}_{\text{train}}$. Add the regression function of your model to the plot.

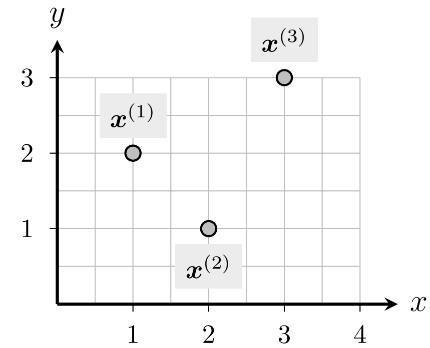

Figure 1: Plot of the training data set $\mathcal{D}_{\text{train}}$.

Question 3 [EX] 2022 (Least squares error)

A colleague of yours provides you with a dedicated test dataset $\mathcal{D}_{\text{test}}$ so that you can thoroughly validate the model you have trained in question 2:

$$\mathcal{D}_{\text{test}} = \big\{(0.5, 3), (1, 1.5), (2.5, 1), (3, 1.5)\big\}.$$

Each tuple has the same form as in question 2. Please answer the following questions:

1. Calculate the least squares error of your model from question 2 on the training dataset $\mathcal{D}_{\text{train}}$. **Hint:** Use $\boldsymbol{\theta}^{\star} = \begin{pmatrix} 1 \\ 0.5 \end{pmatrix}$ in your calculations.

2. Compute the least squares error on $\mathcal{D}_{\text{test}}$! **Hint:** Use $\boldsymbol{\theta}^{\star} = \begin{pmatrix} 1 \\ 0.5 \end{pmatrix}$ in your calculations.

3. Figure 2 visualizes both datasets, $\mathcal{D}_{\text{train}}$ and $\mathcal{D}_{\text{test}}$. How do you rate the model performance given figure 2 as well as the least squares errors you have computed on both datasets? What can you do to improve the model?

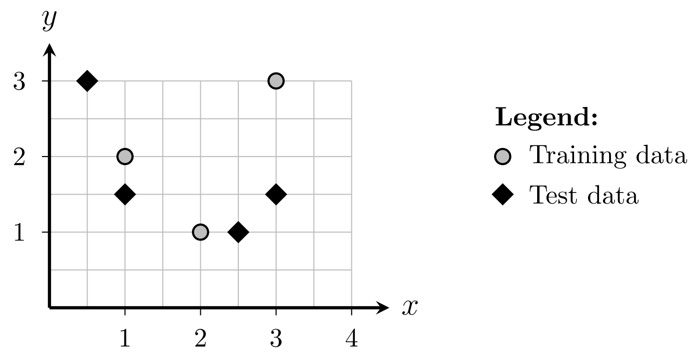

Figure 2: Plot of the complete dataset comprising train and test data.

Question 4 [EX] 2022 (Regularization)

- ☐ Regularization mitigates the danger of underfitting.
- ☐ Regularization mitigates the danger of overfitting.
- ☐ The regularization parameter λ should be chosen as large as possible.
- ☐ The formula for ridge regression is given by $\boldsymbol{\theta}^{\star} = \left(\boldsymbol{X}^{\intercal}\boldsymbol{X} + \lambda \boldsymbol{I}\right)^{-1} \boldsymbol{X}^{\intercal}\boldsymbol{y}$.
- ☐ It is not possible to use regularization and basis functions simultaneously.

Question 5 [EX] 2020, modified (Basis functions)

Name some examples of basis functions. Also do some research online: What other types exist? What is the purpose of basis functions? Can any function serve as a basis function?

Question 6 (Polynomial basis functions)

You are given the following dataset consisting of the two features x_1 and x_2:

Row	x_1	x_2	y
1	3	2	5
2	1	1	4
3	4	3	2
4	1	2	3
5	6	1	1

Please compute the transformations $\boldsymbol{\varphi}(\boldsymbol{x}^{(i)})$ of the data points $\boldsymbol{x}^{(i)}$ using the following polynomial basis functions:

$$\boldsymbol{\varphi} : \mathbb{R}^2 \to \mathbb{R}^6, \qquad \boldsymbol{\varphi}(\boldsymbol{x}) = \begin{pmatrix} \varphi_0(\boldsymbol{x}) \\ \varphi_1(\boldsymbol{x}) \\ \varphi_2(\boldsymbol{x}) \\ \varphi_3(\boldsymbol{x}) \\ \varphi_4(\boldsymbol{x}) \\ \varphi_5(\boldsymbol{x}) \end{pmatrix} = \begin{pmatrix} 1 \\ x_1 \\ x_1 x_2^2 \\ x_1^2 x_2 \\ x_2^3 \\ x_2^5 \end{pmatrix} \in \mathbb{R}^6.$$

Question 7 (Normal equation, polynomial basis functions, and regularization)

The following training dataset is presented to you (see figure 3):

$$\mathcal{D}_{\text{train}} = \big\{(1, 1/2), (2, 1), (3, 1/2), (4, 2)\big\}.$$

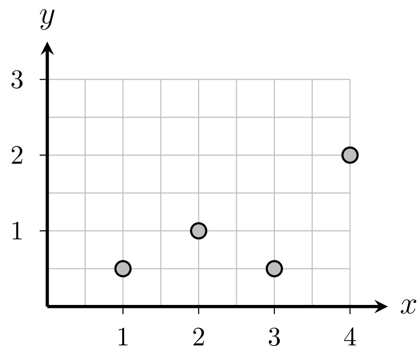

Figure 3: Plot of the training dataset $\mathcal{D}_{\text{train}}$.

Your task is to train a linear regression model on $\mathcal{D}_{\text{train}}$. Due to the fact that the data is slightly non linear, you decide to use **polynomial basis functions** to transform the features in a non-linear fashion. You choose the feature functions

$$\boldsymbol{\varphi}(x) = \begin{pmatrix} 1 \\ x \\ x^3 \end{pmatrix},$$

so that the model function will be of the form

$$h_{\boldsymbol{\theta}}(x) = \theta_0 + \theta_1 x + \theta_2 x^3.$$

In order to prevent overfitting you decide to regularize the model using the regularization parameter $\lambda = 1/2$.

Please answer the following questions:

1. Compute the optimal model parameters $\boldsymbol{\theta}^\star$. *(It is a valuable exercise to compute it by hand, but you may also use software.)*

2. What is the prediction for the unknown data point $x_q = 1/2$?