Artificial Intelligence and Machine Learning

Exercises – 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. Do the following operations result in a matrix, a vector, or a scalar? What are the respective dimensions?

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

Question 2 (Normal equations) ⊛

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 tasks:

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

$$\boldsymbol{\theta}^\star := (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}.$$

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

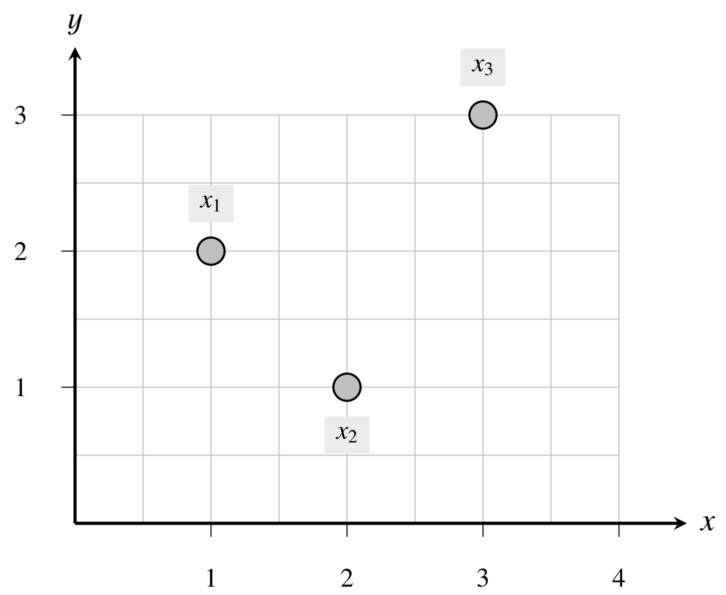

Figure 1: Plot of the training dataset $\mathcal{D}_{\text{train}}$.

Question 3 (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:

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

Each tuple has the same form as in question 2. Please work through the following tasks:

1. Calculate the least squares error of your model from question 2 on the training dataset $\mathcal{D}_{\text{train}}$.

2. Compute the least squares error on $\mathcal{D}_{\text{test}}$.

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?

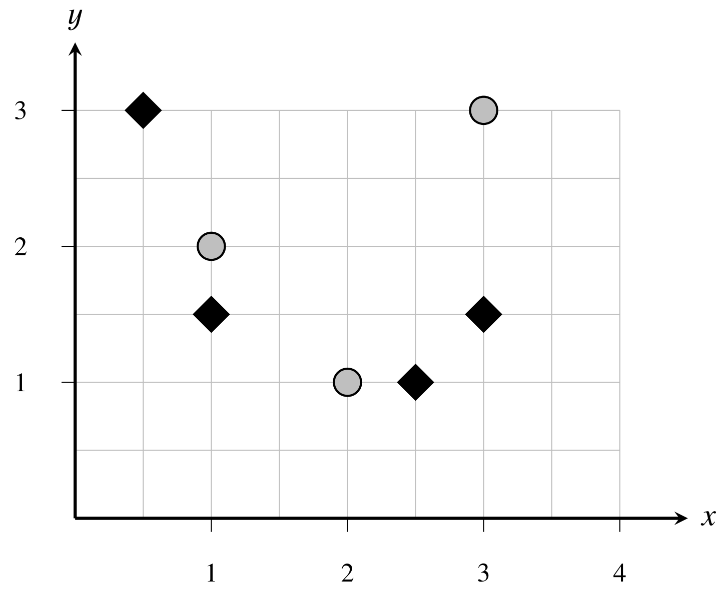

Figure 2: Plot of the complete dataset comprising train and test data. Circles represent elements of the training set and diamonds instances from the test set.

Question 4 (Regularization) ⊛

Tick the correct statements concerning the regularization of linear regression models:

- ☐ 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 $\boldsymbol{\theta}^\star := (\boldsymbol{X}^\top\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$.
- ☐ It is not possible to use regularization and basis functions simultaneously.

Question 5 (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 comprising 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

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

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

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

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

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

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

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

so that the model function will take 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 $\lambda := 1$.

Please answer the following questions:

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

$$\boldsymbol{\theta}^\star := (\boldsymbol{\Phi}^\top \boldsymbol{\Phi} + \lambda \boldsymbol{I})^{-1} \boldsymbol{\Phi}^\top \boldsymbol{y}.$$

2. What is the prediction for the unknown data point $x' := 1/2$?

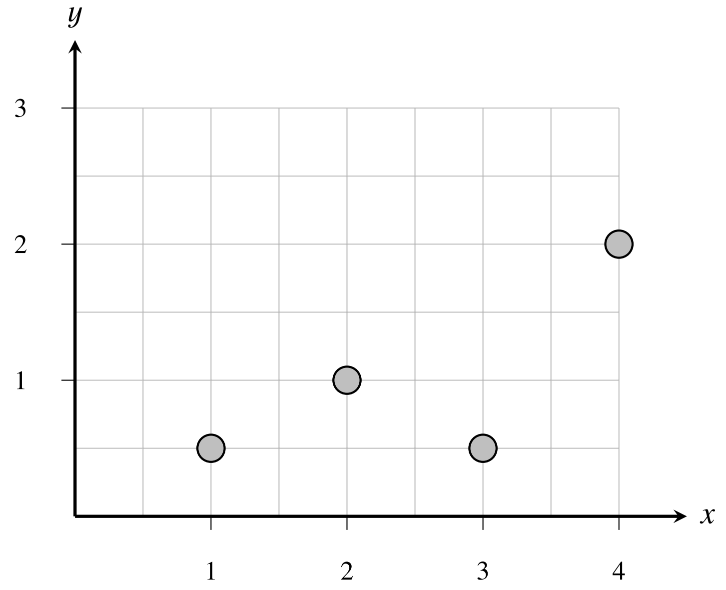

Figure 3: Plot of the training dataset $\mathcal{D}_{\text{train}}$.

Question 8 (Implement linear regression)

Use the following Python snippet to generate a regression problem:

```python
from sklearn.datasets import make_regression

X, y = make_regression(
    100,
    n_features=1,
    noise=35.0,
    random_state=45
)

# uncomment this for task 3
#y = y**2 / 10000
```

1. Solve this regression task **analytically** and plot the results.

2. Now implement the **gradient descent algorithm** to solve the same task. Plot the results and compare them to your analytic solution from task 1.

3. Uncomment the non-linear transformation of the target labels and adapt your code to solve this new task, i. e. implement suitable **basis functions**.

4. Compare the results of your implementation to the results obtained using `sklearn`.