

# Assignment I

## Machine Learning

January 26, 2026

### Abstract

These exercises aim to review the regression problem in supervised learning. Please upload your solutions to Classroom before TBA.

### Problem 1

Consider the dataset from [1]. Make a Python script to solve the regression problem associated with forecasting city and highway fuel consumption using sklearn. You can use two separate linear models for each fuel consumption category. Write down all the assumptions and data preprocessing operations you make.

### Problem 2

Solve problem 1 using the normal equations for linear regression. Compare the solutions to both problems and write down your conclusions.

### Problem 3

Consider the following function:

$$f(x) = 2^{\cos(x^2)}, \quad x \in \mathcal{I} = [-\pi, \pi].$$

The goal is to approximate  $f$  by a polynomial model

$$h(x; \theta, n) = \sum_{j=0}^n \theta_j x^j, \quad \theta = (\theta_0, \theta_1, \dots, \theta_n)^T,$$

for a suitable order  $n$ . Make a Python script to solve the regression problem associated. Notice that the dataset  $D$  consists of a sampling of  $f$  in  $\mathcal{I}$  of size  $m$ . Write down all the assumptions you make. Also, write down the optimization hyperparameters you choose and explain why you picked them that way. Include a plot of the error vs. iterations and a plot of the solution. Don't forget to remark which value of  $n$  you choose and why.

### Problem 4

Solve problem 3 using the polynomial regression model from sklearn [2]. Compare the solutions to both problems and write down your conclusions.

## Problem 5

Solve problem 3 using the normal equations for polynomial regression. Compare the solutions to both problems and write down your conclusions.

## Problem 6

Consider problem 3 but this time using a linear model. Plot the associated learning curves. What do you conclude?

## Problem 7

Consider the following function:

$$f(x) = 2x^2 - 5, \quad x \in \mathcal{I} = [-\pi, \pi].$$

Use polynomial regression with  $n = 20$ . Plot the associated learning curves. What do you conclude?

## Problem 8

Consider problem 7, but this time, use only 5 iterations for training. Plot the associated learning curves. What do you conclude?

## Problem 9

Consider problem 7, but this time, add L2 regularization [3]. Write down your results.

## Problem 10

Consider problem 7, but this time, add L1 regularization [4]. Write down your results and compare them with those from problem 10.

## References

- [1] <https://www.kaggle.com/datasets/rinichristy/2022-fuel-consumption-ratings>
- [2] <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.PolynomialFeatures.html>
- [3] [https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.Ridge.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Ridge.html)
- [4] [https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.Lasso.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Lasso.html)