

# Machine Learning For Data Science I,

## Homework 03

Maj Gaberšček, 27212075

March 2023

### 1 Problem

The problem of this homework was to implement **Ridge (L2)** and **Lasso (L1)** regressions. Then we had to apply the Ridge regression on the superconductivity dataset and find the best regularization weight.

### 2 Ridge regression

Ridge regression is implemented in code as a class **RidgeReg**. When initializing the object, we pass it the argument **weight**.

First method of this class is **fit**, that fits the model to the data and returns nothing. Firstly we append another column of value 1 to learning dataset  $X$ . This is because we want an output  $y(x_1, \dots, x_r) = x_1\beta_1 + \dots + x_r\beta_r + \beta_{r+1}$  (the column of ones will serve to determine  $\beta_{r+1}$ ). Then we use the closed-form solution to calculate vector  $\beta$ .

$$\beta = (X^T X + \lambda I)^{-1} X^T y.$$

Second method of this class is **predict**, that returns predictions of the model on a new dataset, if the model has been fitted before that. We can find predictions, by simply adding a column of ones to the new dataset and then making a dot product with previously calculated  $\beta$  vector.

### 3 Lasso regression

Lasso regression is implemented in code as a class **LassoReg**.

Again, the first method of the model is **fit**. For similar reasons as before, we append another column of ones to the input dataset  $X$ . Then we minimize the function

$$L(\beta) = \sum_{i=1}^n (\beta^T x_i - y_i)^2 + \lambda \sum_{i=1}^k \beta_i^2$$

with the function `minimize` from `scipy` package. We are using the Powell method.

The second method, `predict`, works similarly as in Ridge regression.

## 4 Application

### 4.1 Superconductivity dataset

Firstly, we loaded the dataset and split it into features, test attribute, test target, train attribute and train target values with function `load`.

Then we trained the model for different  $\lambda$  values, from 0 to 3. We saved root mean squared error for each model. Then we chose the weight, that produced the lowest root mean squared error and trained the model with this weight (optimal weight was 0). We then estimated the true root mean squared error by testing the model on last 100 data (that was not used for training). RMSE estimate on testing set was 40.7 (RMSE on training set was expectedly lower, 10.3). Plot of RMSE for different weights (tested on training data) can be seen in Figure 1.

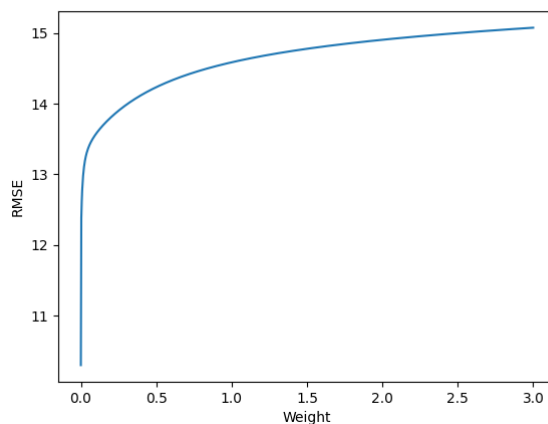


Figure 1: RMSE for different  $\lambda$  (weight) of L2 regression

### 4.2 Tests

At the end, we run some tests for Lasso and Ridge regression for:

- small input dataset (one feature)
- larger training (three features)
- error handling, if predicting on untrained model