Homework 4

April 2022

1 Regression with regularization terms

Solve a simple regression problem using Mean Squared Error (MSE) loss with three different penalty or regularization terms. The Forest Fires dataset (forestfires.csv) has 13 columns where the first to the 12-th columns are attributes (Please see forestfires.txt for an introduction to these attributes). The last column is the burned area of the forest fire. Given a sample $X \in \mathbf{R}^{12}$, you need to find a function f to predict the burned area of forest fires $y \in \mathbf{R}$. The MSE loss is given as

$$Loss_{MSE} = \frac{1}{N}\sum_{i=1}^{n}\left(y_i - f(X_i)\right)^2, \tag{1}$$

where X_i denotes the i-th training sample and y_i denotes the burned area of the forest fire. N denotes the number of training samples.

Specifically, the dataset (forestfires.csv) has 517 samples. You can use the first 450 samples as the training set, and the last 67 samples as the test set.

Here are four methods you need to implement:

(1) Find a linear model to solve the regression problem. In other words, the linear model is $f(X_i) = W^T X_i$, where $W \in \mathbf{R}^{12}$. To this end, you need to compute the MSE loss and use gradient descent to find a weight matrix W.

(2) Using Ridge Regression to solve the regression problem. The ridge regression you need to implement is on page 3 of lecture 3.

(3) Using RBF kernel regression to solve the regression problem. The kernel regression you need to implement is on page 5 of lecture 3, and the RBF kernel is on page 12 of lecture 2.

(4) Using Lasso Regression to solve the regression problem. The lasso regression you need to implement is on page 8 of lecture 3.

Please provide the following results for four methods:

(1) Please report the total prediction error (*i.e.* $Error_{test} = \sum_{i=1}^{m} (y_i - f(X_i))^2$, where m denotes the number of the testing samples) on the test set using four methods. In particular, you should try different penalty coefficients λ in methods (2)-(4), and give the prediction errors of at least two penalty coefficients λ for each method in methods (2)-(4).

(2) Please visualize the weight vector W for each method. Specifically, given a weight vector $W = [w_1, w_2, ..., w_{12}]^T \in \mathbf{R}^{12}$, you first need to sort $\{w_1, w_2, ..., w_{12}\}$ according to their absolute value from largest to smallest. Then, you need to visualize the new sorted absolute weight vector $\{|w_1|, |w_2|, ..., |w_{12}|\}$. For example, given a 12-dimensional weight vector $W = [-2.5, 4.1, 2.8, -3.4, 0.5, 0.45, 1, 1, 10, 0.8, 0.9, 6]^T$, we can get its sorted absolute weight vector $[10, 6, 4.1, 3.4, 2.8, 2.5, 1, 1, 0.9, 0.8, 0.5, 0.45]^T$. Finally, we visualize the sorted absolute weight vector in Figure. 1. Please draw distributions of weights of four methods in an image for comparison.

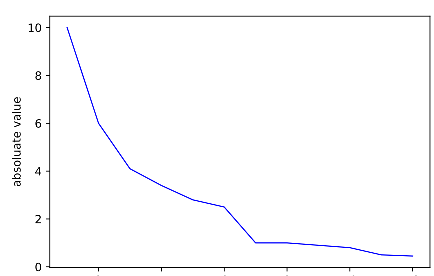

Figure 1: Visualization of the absolute value of the weights.

(3) Please provide some analysis and discussion about your experimental results.

More information about the dataset can be referred to as follows:
https://archive.ics.uci.edu/ml/datasets/Forest+Fires