

Machine Learning Course 2024 Spring: Experiment 1

March 12, 2024

1 Overview

In this experiment, you will implement linear models, including linear regression, ridge regression, and logistic regression and evaluate trained models on test set. You will also compare your implementation with the models in sklearn.

2 Dataset Description

The datasets are ready for you and are stored in the file *'diabetes_train.csv'*, *'diabetes_test.csv'*, *'pendigit_train.csv'*, *'pendigit_test.csv'*. Each row in the files is a sample, with the last column being the label and the other columns being the features.

3 Task Description

3.1 Task 1

Given the *diabetes* dataset, implement the **linear regression** model with **gradient descent**, and evaluate the trained model on the test set.

3.2 Task 2

Given the *diabetes* dataset, implement the **ridge regression** model with **gradient descent**, and evaluate the trained model on the test set. **5-fold cross-validation** will be employed to select the proper value of the hyperparameter, and the influence of these values on the trained model will be evaluated.

3.3 Task 3

Given the *pendigit* dataset, implement the **logistic regression** model with **gradient descent**, and evaluate the trained model on the test set.

3.4 Task 4

Compare the trained models with standard models in **sklearn**. Discuss the difference between them and try to analyze the possible reasons.

4 Requirements

- Implement learning algorithms with **Python**. Basic data analysis libraries such as **Numpy**, **Pandas**, **Matplotlib**, etc., are allowed to use. However, you can't use highly integrated libraries such as **sklearn**, **PyTorch**, **Tensorflow**, etc.
- Record the procedure of experiments in **Jupyter Notebook**. It is recommended to use **Markdown** in Notebook to properly explain the procedure.
- Your submission includes:
 - source code: python notebook (.ipynb);
 - 'report.doc' file : your report. Refer to the report template for the format.