

Machine Learning and Intelligent Systems

Project 3: Ridge Regression

REPORT

Ridge Regression Implementation

1. **Initialization:** The RidgeRegression class is initialized with a regularization parameter alpha (lambda), and the weight vector W is set to None.
2. **Adding Bias Term:** A column of ones is added to the feature matrix X to account for the bias term in the linear model.
3. **Analytical Solution:** The weights are calculated analytically using the formula:

$$W = (X^T X + \alpha I)^{-1} X^T y$$

where I is the identity matrix with the same dimension as $X^T X$. The first element of the regularization term is set to zero to avoid regularizing the bias term.

4. **Prediction:** The predict method computes the dot product of the feature matrix X (with the bias term) and the weight vector W .
5. **Performance Metric:** The Mean Squared Error (MSE) is used as a performance metric to evaluate the model.

Performance

On the 100m Olympics Dataset : The dataset consists of the years and corresponding finishing times of the 100m sprint event in the Olympics. The Ridge Regression model is trained on this dataset to predict the finishing times based on the year. Computing the MSE after choosing the best lambda, we can see that the MSE is 0.047, which is quite low. We also did the performance evaluation using a separate test dataset to ensure there was no overfitting.

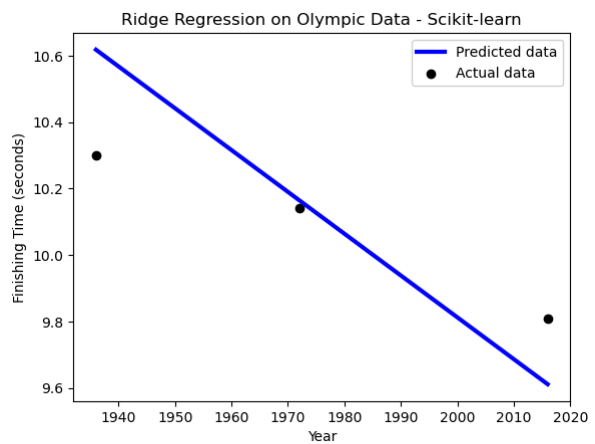
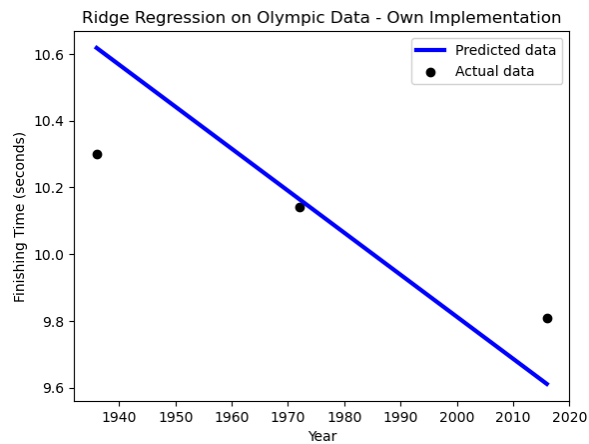
On the California Housing Dataset : The implementation was also tested and confirmed to be working with a different dataset.

Choice of Lambda

The best lambda was found using hyperparameter tuning, where the dataset is split into training, validation and test sets, and the model is trained with different lambda values. The lambda that results in the lowest validation MSE is chosen.

Comparison with Scikit-Learn's Implementation

In comparison with the performance of the same dataset in Scikit's Ridge, we found that the MSE was 0.047, same as the MSE we got with our implementation.



GPT use

GPT was used to debug and add comments to the base code that was provided. It was also used to create the structure of the report and after the report was written, to check for any grammatical errors. It was also used to generate the code for plotting data.