(3) Write a linear regression algorithm by yourself.

- Use Boston House Dataset (506 samples and 13 feature variables),

predict the value of prices of the house using the given features.

- Split input data into training and testing sets (the testing set includes 10%ofthe samples).

- The output of the algorithm should include the model learned from the whole training set, the average mean squared error (MSE) on training set via 10-fold cross validation.

For ridge regression model, determine the value of hyper-parameter via 10-fold cross validation.

(4) Write the experimental process and results into an experimental report. (strict content, abundant pictures and words).

- See if there are any missing values in the data.

Missing Values:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CRIM** | **ZN** | **INDUS** | **CHAS** | **NOX** | **RM** | **AGE** | **DIS** | **RAD** | **TAX** | **PIRATIO** | **B** | **LSTAT** | **MEDV** |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

- Exploratory data analysis (e.g., plot the distribution of the target variable

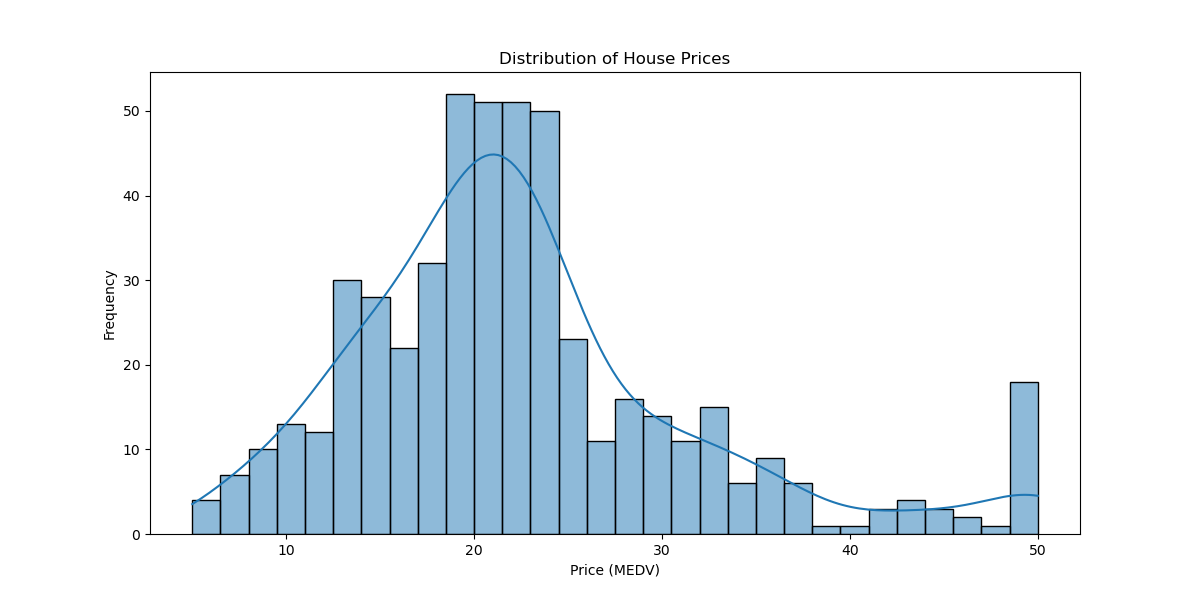


Figure 1 Distribution of House Prices

, create a correlation matrix that measures the linear relationships between the variables and plot the correlation matrix

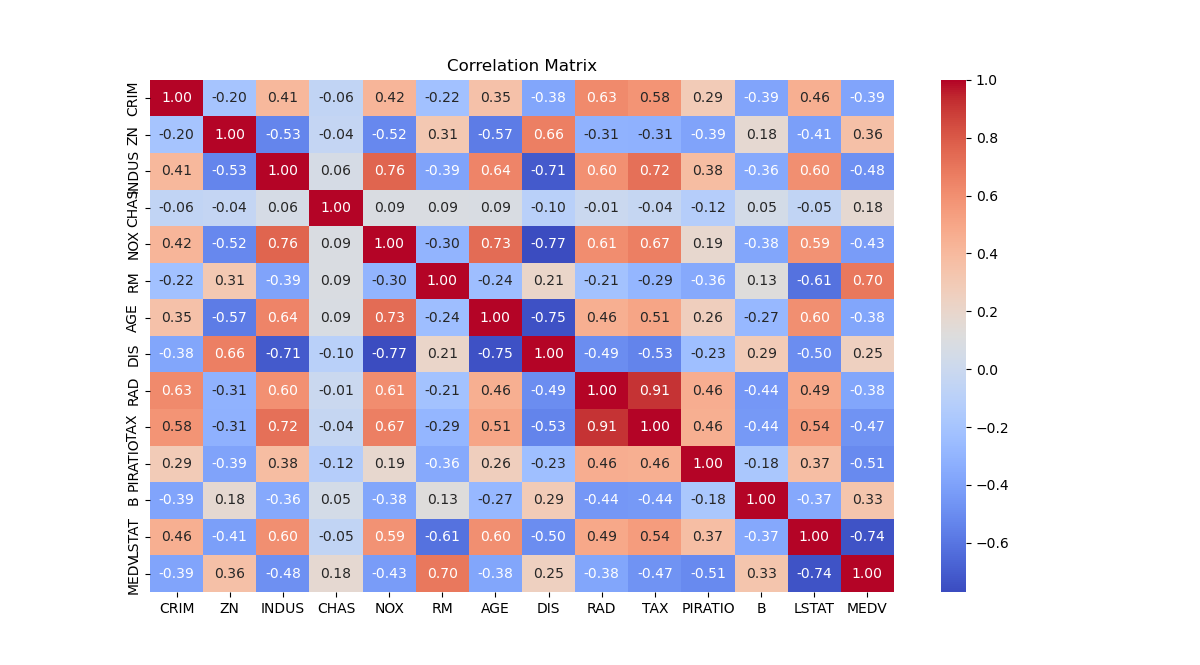


Figure 2 Correlation matrix between the variables

, show the meansquared error (MSE) on testing set, plot a scatter plot between the original house price and predicted house prices).

1. Linear Regression Model:

- Model Coefficients: [-1.19886262e-01 3.99134691e-02 2.12938504e-02 2.77565167e+00

-1.85854960e+01 3.75579160e+00 4.57076424e-03 -1.47064595e+00

3.11878023e-01 -1.18109903e-02 -9.47556337e-01 1.03287982e-02

-5.50096256e-01]

- Average MSE on Training Set (10-fold Cross Validation): 24.907888998639386

2. Ridge Regression Model:

- Best Hyper-parameter (alpha) selected: 0.1

- Average MSE on Training Set (10-fold Cross Validation): 24.909414528624904

3. Evaluation on Testing Set:

- Linear Regression Model MSE on Testing Set: 14.995852876582926

- Ridge Regression Model MSE on Testing Set: 14.933506462681148

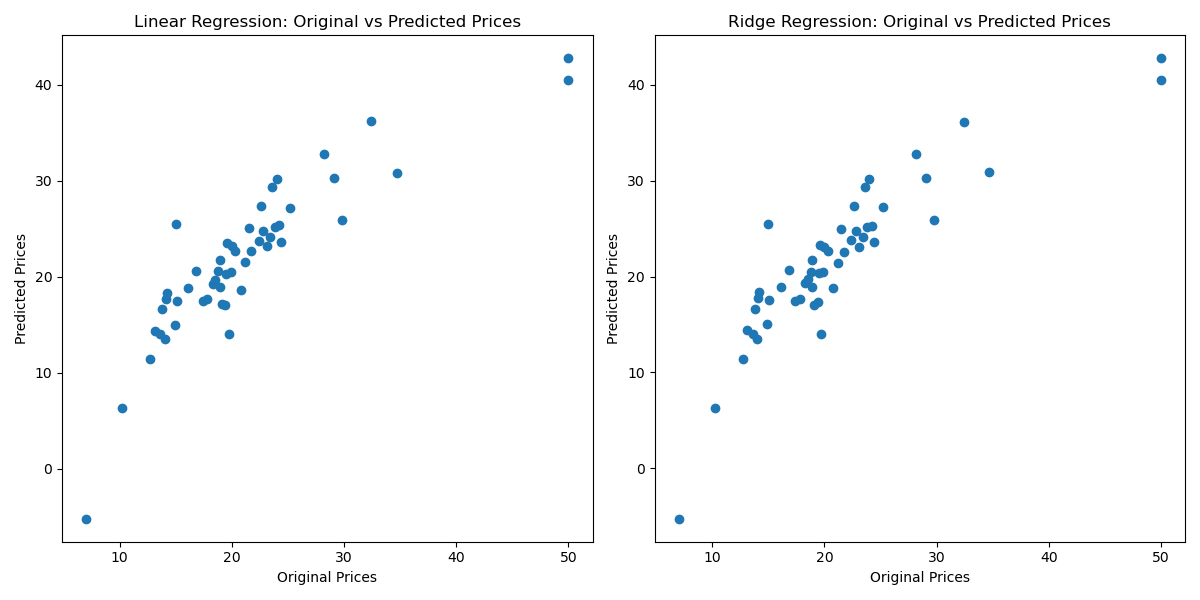


Figure 3 Original house price and Predicted house prices