

Exercise 3 - Regression

November 18, 2019



1 Linear Regression

a) Ordinary Least Squares (5 points)

Implement an ordinary least squares regression model using numpy/pandas to predict the value of a house in Boston using the Boston Housing Dataset in the folder *data* (file *bostonhousing-dataset.csv*). The target variable is in the last column, it is called *medv* (median value of owner-occupied homes in \$1000s). You can find an explanation of each attribute in the dataset on Kaggle ([link](#)). Evaluate and report the mean squared error (MSE) of your model on the dataset.

b) Basis Function Features (3 points)

Compute polynomial or radial basis function features from the source data. Optimize the basis functions for the task of predicting the value of a house (i. e. tune the degree of the polynomials or means and scale of the radial basis functions). Which basis functions worked best?

c) Regularization (2 points)

Explain in your own words what *regularization* is, why it is beneficial and which kinds of regularization you could apply to a linear regression model. Explain what *ridge regression* is.

d) Bonus Question 1 (1 point)

Which are the three most important features for the house value prediction according to your linear regression model? Why?

e) Bonus Question 2 (1 point)

Plot the residuals for your regression model (y-axis) and the predicted values (x-axis). What can the residuals tell you about the performance of your model?