

Fahim Khan
CS6923 Machine Learning
5/1/20

Final Project: Lasso Regression

Lasso Regression is the extension for Linear Regression.

Lasso Regression is a regularization method on linear regression, where non-essential features become or get closer 0 in w .

- The numpy and sklearn versions both use coordinate descent
- I did coordinate descent similarly to lectures, while sklearn does it somewhat differently
 - So the numpy version converges to 0 much faster, therefore a less number of iterations may be needed by the numpy version compared to sklearn's version

Lasso regression, both numpy and sklearn, had slightly worse training accuracy but better test prediction accuracy on all test data. This is to be expected since Lasso regularization helps reduce overfitting which makes predictions more accurate than regular linear regression.

The mean squared error (MSE) and coefficient of determination (R^2) are used to score how well Lasso Regression does compared to Linear Regression.

K-fold validation is used in the numpy version to compare how well numpy Lasso did to the original HW's Linear implementation.

The digits dataset was a good Sklearn dataset to use since the MSE was similar to Boston's dataset, and no additional preprocessing was needed.

The diabetes dataset was another choice and was recommended for regression, but the MSE was very high across both Sklearn and numpy, even when the data had a fixed intercept and went through polynomial transformations.

linear regression train score: 0.5979800626243363	lasso regression train score: 0.595761068107747
linear regression test score: 0.5740460038277273	lasso regression test score: 0.5753167560957702
linear MSE test: 3.452475459530763	lasso MSE test: 3.4421756594115345

Sklearn linear regression results

Sklearn lasso regression results ($\lambda = 0.01$)

linear regression train score: 0.5820634460338829	lasso regression train score: 0.5807017771843322
linear regression test score: 0.5644507133517508	lasso regression test score: 0.5669872636401383
MSE_test = 3.53 MSE_train = 3.438	MSE_test = 3.51 MSE_train = 3.449

Numpy linear regression results

Numpy lasso regression results ($\lambda = 0.01$)