Mia Rodgers: <https://github.com/miamrodgers/4310-ML/blob/main/ME6_regression/regression.ipynb>

Alex Larsen: <https://github.com/alarsen123/ML-HW/blob/main/ME6_regression/ME6_regression/regression.ipynb>

ME6

In this exercise, we practiced linear regression with gradient descent on the sklearn breast cancer dataset. We first used the Normal equation to obtain the best theta vector which was:

[ 3.02181174e+00 2.17772056e-01 -4.54546867e-03 -2.37398610e-02 -3.17834750e-04 -8.46891370e-02 4.22203525e+00 -1.39799728e+00 -2.14183303e+00 -1.02709200e-01 -3.32616097e-02 -4.34955932e-01 6.75847233e-03 2.25202577e-02 9.23217886e-04 -1.58543207e+01 -6.49034090e-02 3.56546799e+00 -1.05679513e+01 -1.69734069e+00 7.14644016e+00 -1.95183121e-01 -7.15937520e-03 2.43505057e-03 1.01122332e-03 -5.42856861e-01 -6.71582941e-02 -3.81191215e-01 -4.64309895e-01 -5.56787546e-01 -4.30348309e+00]

Then we did the same using gradient descent where the best theta was a 31 by 569 array so it is too big to put in this summary. Next, we tested six different modeling techniques which included: (1) linear regression without regularization, (2) ridge regression, (3) ridge regression with feature normalization, (4) Ridge regression with regularization parameter: alpha, (5) Lasso regression, and (6) Lasso regression with regularization parameter: alpha.

A picture containing shape

Description automatically generatedModel performance:

1. R-squared score (training): 0.782   
   R-squared score (test): 0.729   
   RMSE: 0.251

A picture containing shape

Description automatically generated

1. R-squared score (training): 0.744   
   R-squared score (test): 0.717   
   RMSE: 0.257
2. A picture containing shape

   Description automatically generatedR-squared score (training): 0.741   
   R-squared score (test): 0.721  
   RMSE: 0.255
3. Graphical user interface

   Description automatically generated with medium confidence

Chart

Description automatically generated

1. R-squared score (training): 0.690   
   R-squared score (test): 0.688  
   RMSE: 0.270
2. A picture containing graphical user interface

   Description automatically generated

Out of all of these models, it seems like ridge regression and lasso regression seem to have the best performance.