Thomas Chung
Fang
12 February 2018

Homework 3

Directions:

Perform linear regression directly using the closed form solution. Compute the RMSE value on the training data and test data, respectively.

Perform ridge regression directly using the closed form solution. Use k-fold cross validate (k=5) to select the optimal λ parameter. Compute the RMSE value on the test data.

You can begin by running the solver with λ = 400. Then, cut λ down by a factor of 2 and run again. Continue the process of cutting λ by a factor of 2 until you have models for 10 values of λ in total.

Perform linear regression using the gradient descent algorithm. Compute the RMSE value on the training data and test data, respectively.
For the initial weights, you can just use Gaussian N(0, 1) random variables. Define "converging" as the change in any coefficient between one iteration and the next is no larger than 10−5.

Perform ridge regression using the gradient descent algorithm. Compute the RMSE value on the test data.

See Code.

Output:

The following output is what happens when I run my implementation of linear regression, ridge regression, and the gradient descent versions of the regression models.

Linear Regression RSME (Training): 0.12768967421762195
Linear Regression RSME (Test): 0.14583464490949063
Linear Regression Gradient Descent RSME (Training): 0.43129063925162553
Linear Regression Gradient Descent RSME (Test): 0.4503160262381095
Ridge Regression RSME (Training): 0.12879701459879794
Ridge Regression RSME (Test): 0.14574650707058057
Ridge Regression Gradient Descent RSME (Training): 0.30372371071636844
Ridge Regression Gradient Descent RSME (Test): 0.34292600415531543

This output was achieved by initializing my gradient descent models with a weight vector of *Gaussian N(0,1)*. We see from the training RSME values of our regression models that we have RSME values > 0. This indicates to us that our regression models are not overfitting the data set.

The **next output** is what happens when we lower the tolerance 'epsilon' (from 1e-5 to 1e-7) of our gradient descent models in an attempt to achieve a better RSME value.

Linear Regression RSME (Training): 0.12768967421762195
Linear Regression RSME (Test): 0.14583464490949063
Linear Regression Gradient Descent RSME (Training): 0.14262873888301372
Linear Regression Gradient Descent RSME (Test): 0.16074912459241156
Ridge Regression RSME (Training): 0.12879701459879794
Ridge Regression RSME (Test): 0.14574650707058057
Ridge Regression Gradient Descent RSME (Training): 0.13036952269663335
Ridge Regression Gradient Descent RSME (Test): 0.14625334291806413

And the next **final output** is what happens when we lower the tolerance 'epsilon' (from 1e-5 to 1e-8):

Linear Regression RSME (Training): 0.12768967421762195
Linear Regression RSME (Test): 0.14583464490949063
Linear Regression Gradient Descent RSME (Training): 0.1316184934671395
Linear Regression Gradient Descent RSME (Test): 0.14985952942141217
Ridge Regression RSME (Training): 0.12879701459879794
Ridge Regression RSME (Test): 0.14574650707058057
Ridge Regression Gradient Descent RSME (Training): 0.12908269401291708
Ridge Regression Gradient Descent RSME (Test): 0.14491655322928795

As we can see, modifying our tolerance parameter allowed our gradient model to approach the RSME value of the close form solutions, but at the cost of our processing speed.