

Machine Learning and Computational Statistics, Spring 2016

Homework 2: Lasso

Due: Monday, February 15, 2016, at 6pm (Submit via NYU Classes)

Instructions: Your answers to the questions below, including plots and mathematical work, should be submitted as a single PDF file. You may include your code inline or submit it as a separate file. You may either scan hand-written work or, preferably, write your answers using software that typesets mathematics (e.g. \LaTeX , \LyX , or MathJax via iPython).

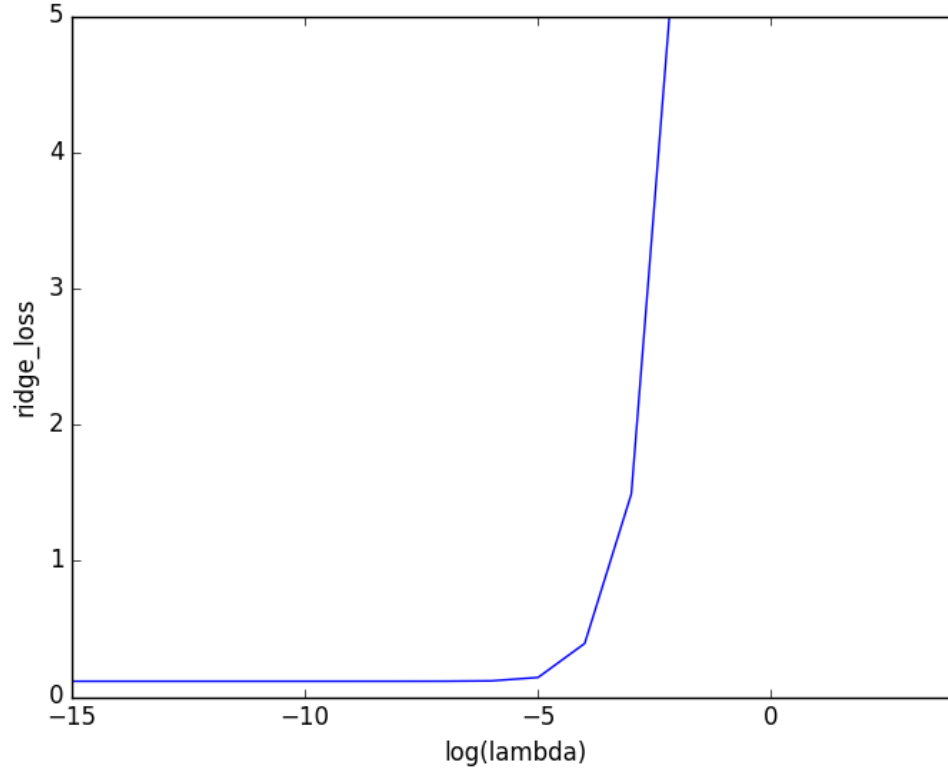
1 Preliminaries

1.1 Dataset construction

Done as part of `createdataset` function in the code that is submitted along.

1.2 Experiments with Ridge Regression

1. The λ value chosen is 10^{*-10} . It can be seen from the graph that the validation loss is minimum at around that value of λ .



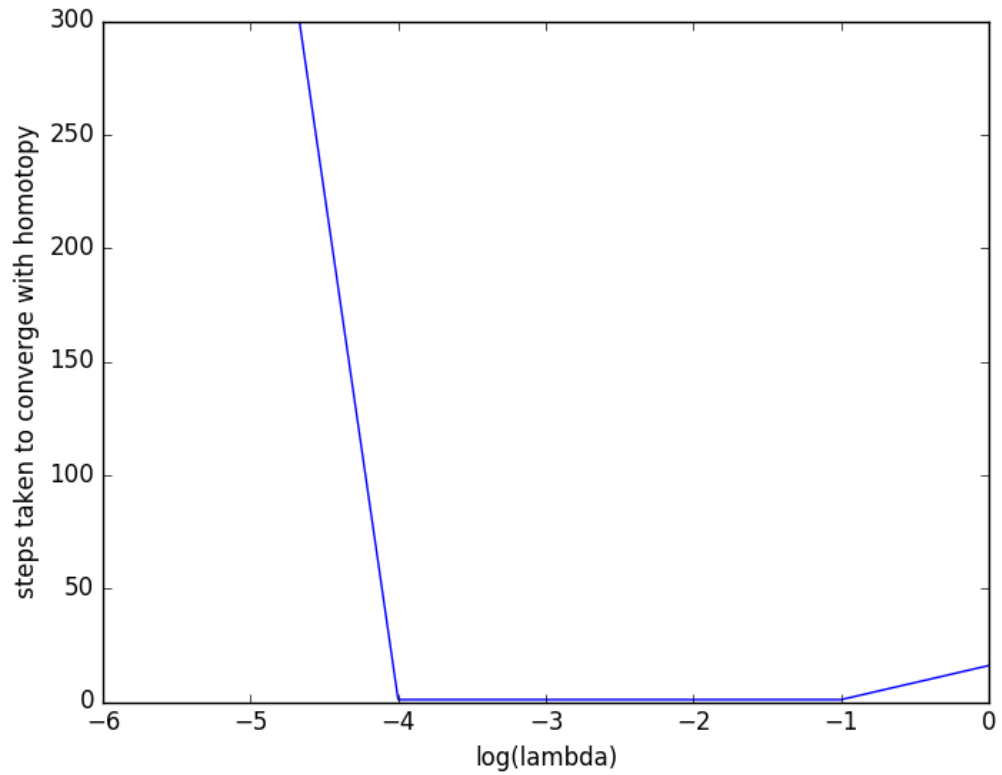
The value of θ where loss converges for $\lambda = 10^{*-10}$ is given below

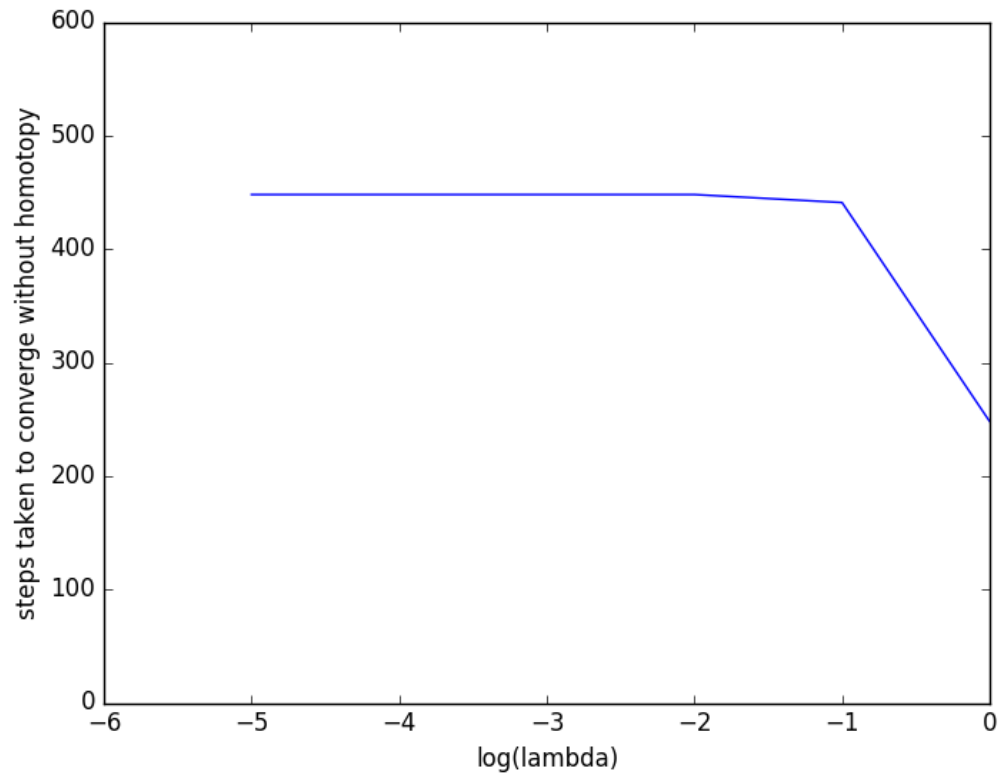
```
[ -9.97078811e+00 -9.94896606e+00 -9.91693187e+00 -9.75743181e+00 9.85992999e+00 1.01373407e+01
-9.79856957e+00 1.01281901e+01 -9.80381646e+00 1.00731125e+01 2.72097142e-01 6.48934430e-
02 -2.78305362e-01 -1.80346366e-01 4.34798615e-02 1.42873962e-01 -4.65720912e-02 2.83894078e-
02 3.39061435e-02 1.41480673e-01 2.29230737e-01 -1.30172965e-01 4.73448622e-02 1.40895080e-
01 -3.04825570e-01 -1.81369197e-01 -1.63071622e-01 2.55285980e-01 -8.64921547e-02 -1.48544968e-
01 -3.85508105e-01 -1.87926893e-02 -2.43938673e-01 4.44195422e-01 -2.76351680e-01 9.36075273e-
02 -4.97978168e-03 1.80403702e-01 -1.28208184e-01 -1.47787388e-01 -1.81369987e-01 -3.94292494e-
02 -1.19387104e-01 -1.14322493e-01 -1.56063159e-01 1.27033886e-01 -1.62222500e-01 -2.53772872e-
02 1.49366859e-01 -5.28729829e-02 -9.93249974e-02 -5.73802138e-02 -1.25475286e-01 3.05958118e-
01 -5.14486295e-02 2.99610761e-02 1.75485414e-01 -2.44970598e-01 -1.38078673e-01 -5.43552090e-
03 3.43687433e-02 1.22835705e-01 -3.21331952e-01 5.71236960e-03 -1.73771826e-02 2.05816282e-
02 -2.43973139e-01 4.30598523e-02 3.01921036e-01 -9.63561659e-02 2.08044204e-01 5.88348587e-
03 -5.48503830e-02 3.62098325e-01 3.16021485e-02]
```

Here, there are no components with true value=0. However, even if we consider a threshold of $10e-3$ zero coefficients remain at zero count.

generated our data set as zero is zero in the solution as well. This confirms the spars nature of lasso solution.

3. The time taken by homotopy solution is 0.902 units and without homotopy for same set of λ the time taken is 3.851 units. Homotopy solution gives faster results. This can be seen in the form of graphs also. In homotopy solution, for first value of λ the steps taken is same as that of without homotopy. However for next set of λ , for homotopy solution the steps taken is almost close to 1 and for without homotopy steps remain almost same.





4. Lasso regularization without vectorization took 3.96 units of time. Lasso regularization with vectorization took 0.24 units of time. This shows the power of vectorization.

Vectorization form:

$$a_j = 2X^T X[j, j]$$

$$c_j = 2X[, j](y - Xw + w_j X[, j])$$