

## Programming assignment 2

Implement householders and back solving for the following:

1. Your datasets
  1. The test dataset provided
    1. Four different error rates (noise in the target value)
    2. One is labeled in the name (0.05) of the file
    3. The others are mysteries until you find them.
    4. Your predictor should be able to solve all of them.
  2. The abalone dataset
  3. You will be using your sampler you created to make a test set and a training set
2. Goals for class
  1. Learn how to go from pseudo code to machine learning algorithm
  2. Learn how to evaluate your algorithms using datasets with known errors
  3. Get the best result possible on the abalone dataset using the algorithm you created
  4. You should come out with an understanding of the parts and details of solving over fit problems.
3. Logistics:
  1. The programming assignment will be due by the end of class on Monday, May 9th
  2. For each dataset provide the RME (root mean square error).
  3. You may work with one other person on the programming projects. Go to gradescope and register your partner (or say you are working alone) along with the programming language you have selected
    1. Java, Python or Julia

Problem Description:

1. Householder's
  1. We covered householder's method in class
  2. I have provided some sample code for implementing a single householder's reflection in `cse151/src/test/java/com/mangione/cse151/householders/HouseHoldersTest.java`
  3. This code is for a single pass
    1. Take the output of this process and then take the sub matrix (start with row 1 ,col 1) and run again
    2. Repeat until you have (m-1) smaller and smaller matrices
    3. Reassemble the all of the pieces into a matrix of size m x m
    4. This matrix should be an upper triangular matrix ready for back solving
    5. Refer to lecture 11 if you have questions
2. Back solving

1. The algorithm and pseudo code are provided in lecture 12
2. The result will be
  1. A weights vector (column 0 of your matrix)
  2. An offset vector
3. Evaluating your algorithm
  1. Take the weights and offset and make predictions and compare against the targets in your test set
  2. See lecture 12 for a description  $Y = mx + B$  (where B is your offset vector)
  3. Calculate the RMSE (lecture 12) for each dataset.

Some helpful hints for the assignment

1. Design techniques will be very helpful because you will be running the same set of algorithms, outputting the results for many runs on many files
2. Please have fun.
  1. The algorithm should be straight forward
  2. This is more about getting all of the techniques in place in order to run training algorithms
3. You are being given two weekends to do this assignment. Make sure you start soon so that we can answer any questions you have
4. This is more about putting it all together than anything else and, as such, each group will take a different path to understanding