CS 6375
ASSIGNMENT ___1_____

Names of students in your group:1

Number of free late days used: ____0_____

Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

scikitlearn/numpy/pandas documentation
hints video
piazza

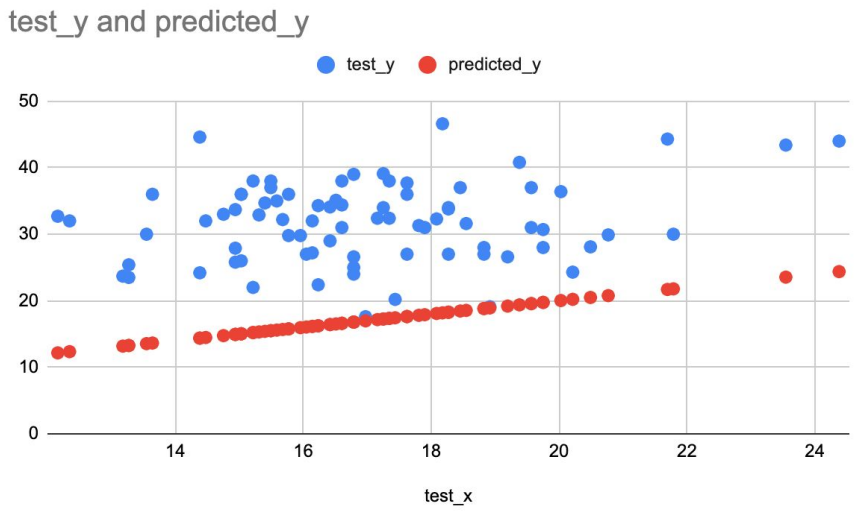
n=100, learningrate=0.001
Result: 1.6109065676120495 0.9255790154372301
RMSE: 16.009547905479405

n=5000, learningrate=0.00001
Result: 1.0028738781790754, 0.9530058198924825
RMSE: 16.15608576743557

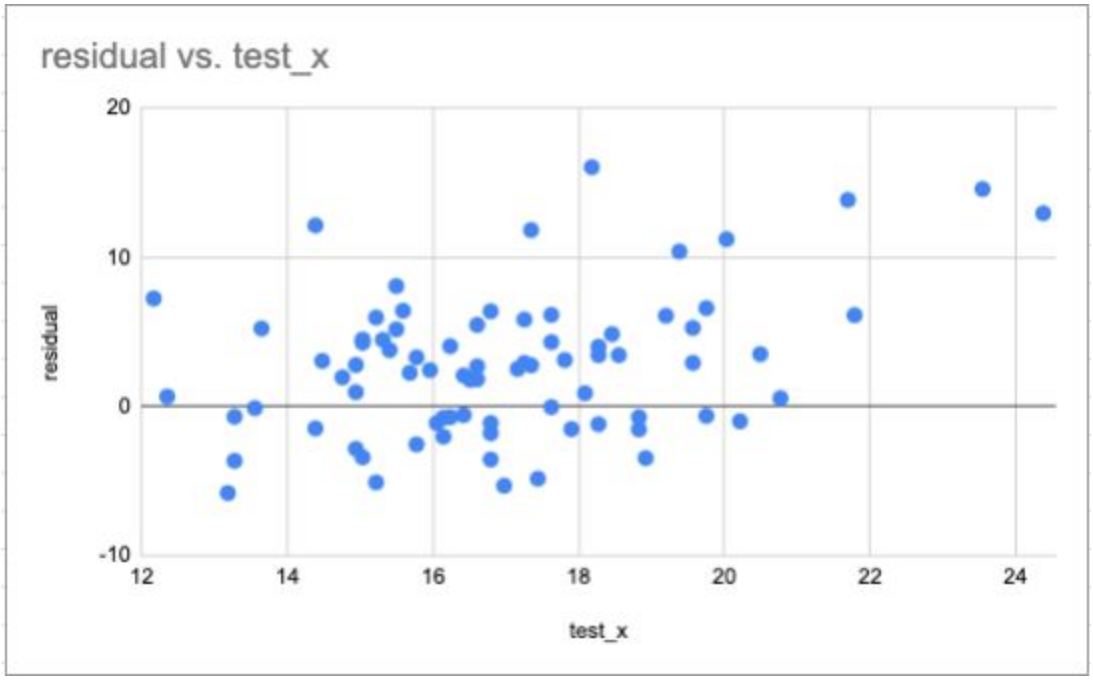
n=1000, learningrate= 0.0001
Result: 1.038803216175747, 0.951322190053528
RMSE: 16.148374100008414

6. Am I satisfied I found the best solution
A: No, The RMSE isn't too bad compared to the one from part 2. Part 2 gave me around 10, here I got 16. But when I look at the plot, the line is just clearly low. It is probably because I started my intercept estimate low and it settled in some local optimum instead of raising itself to the absolute best value. I don't think this can be extended to other datasets either.When I tried my algorithms approach for multiple predictor variables, I got overflow errors.

Plots:



This plot is from part 1



This plot is from part 2