Assignment 2 - Report

Task 1:

In task one we were asked to fit decision stumps i.e. decision tree with a single split. For which I obtained the following results

Best attribute = rm

Best S value = 6.8

Train MSE = 48.0096821937941

Test MSE = 46.525045412662

Surprisingly Test MSE is lower than the Train MSE.

Task 2:

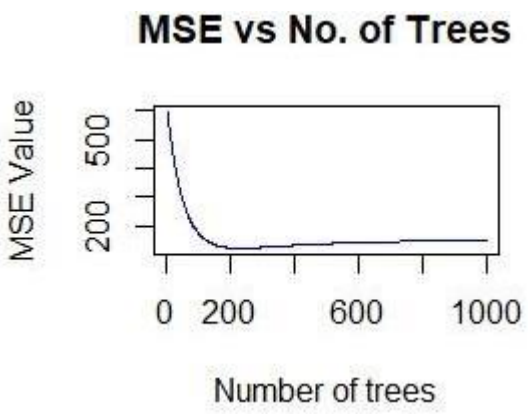
We were asked to implement Boosted Decision Stumps with 1000 trees and 0.01 as our learning rate.

My Test MSE for the same came out to be 148.374779449384 which is abnormal.

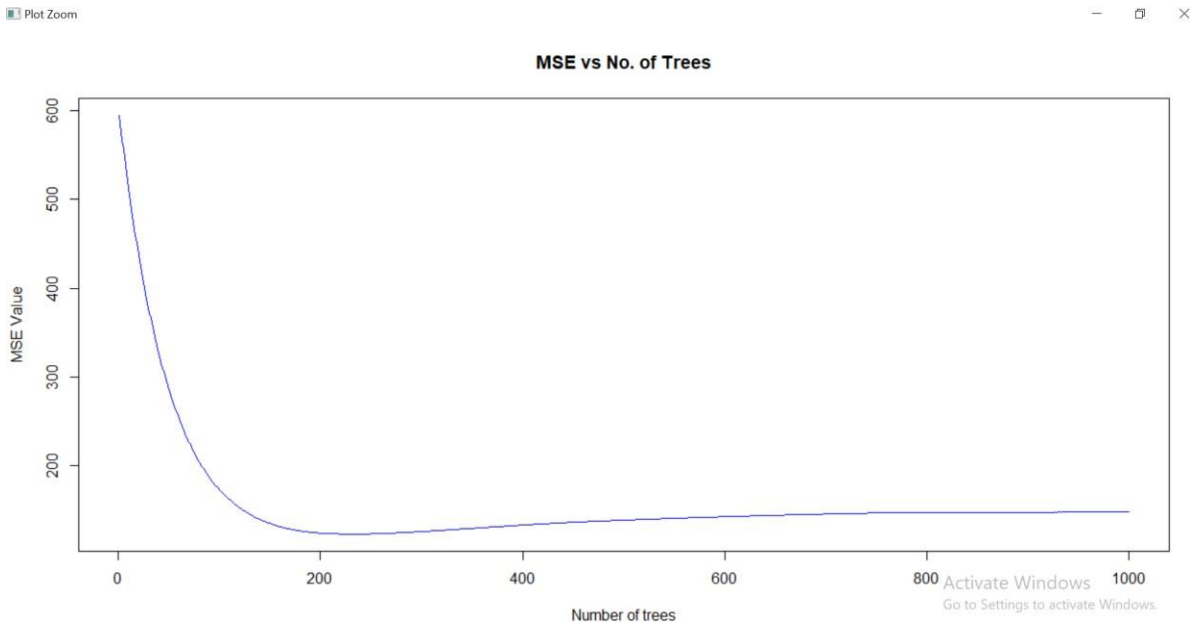
I tried but I got the same value over and over. I don't understand my mistake in the code.

Task 3:

For this task we were asked to find MSE's for 1000 trees and plot the same.



Zoomed version



We see lowering of MSE at high margin between the range of 0 to 175 trees. Then it becomes pretty much steady and after getting to the lowest MSE at around 200 trees it starts to increase again. Hence, we can say it is overfitting.