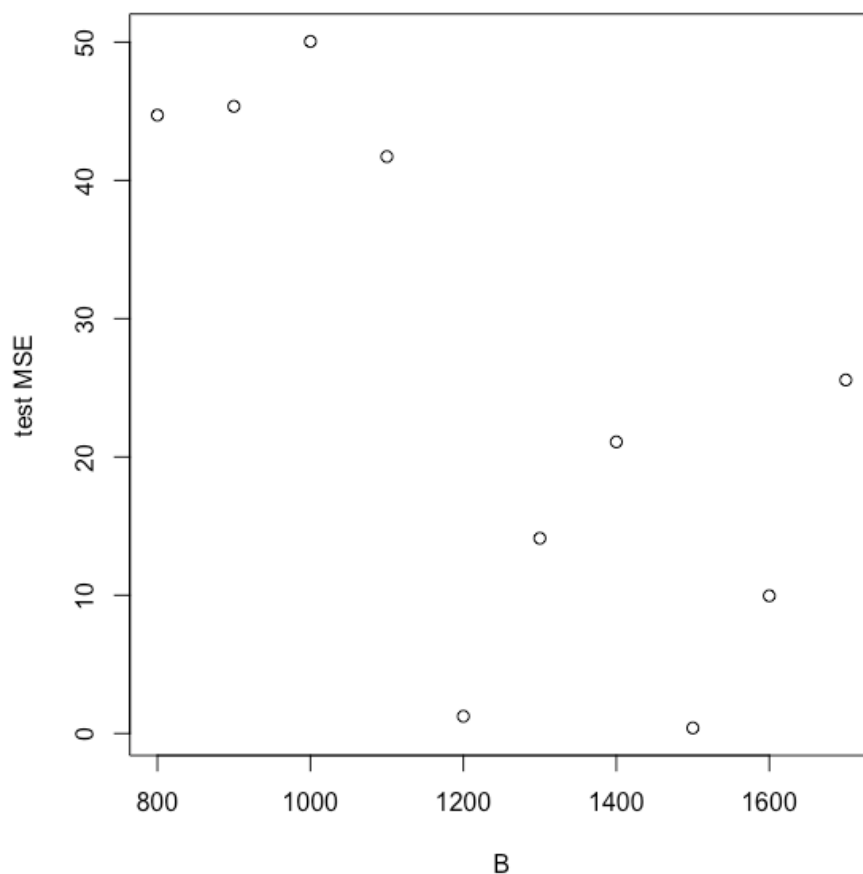


Data Analysis Coursework Assignment 2

Boosting Decision Stump

1. Train your DS implementation on the training set. Find the MSE on the test set. Include it in your report.
5.166631
2. Train your BDS implementation on the training set for learning rate = 0.01 and $B = 1000$ trees. Find the MSE on the test set. Include it in your report.
50.04609
3. Plot the test MSE for a fixed value of learning rate as a function of B (the number of trees) for as large B as possible. Do you observe overfitting? Include the plot and answer in your report.



In this plot, minimum test MSE is when $B=1500$, as 0.3973866. As this plot shows, in this range ($B=800$, $B=1700$), boosting doesn't overfit exactly. In fact, observation demonstrates that test MSE hits almost zero test MSE given $B=1500$. It seems like the algorithm could learn from learning rate arbitrarily without overfitting.