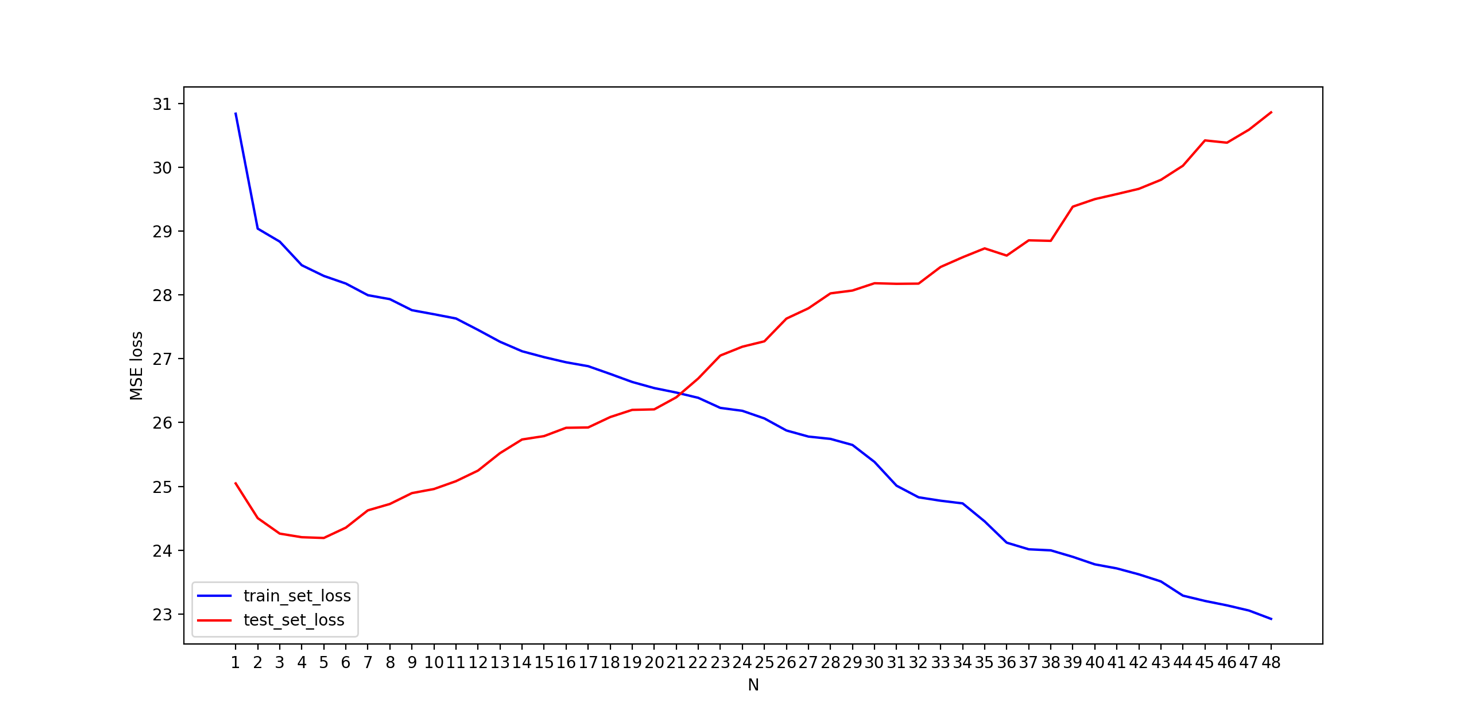
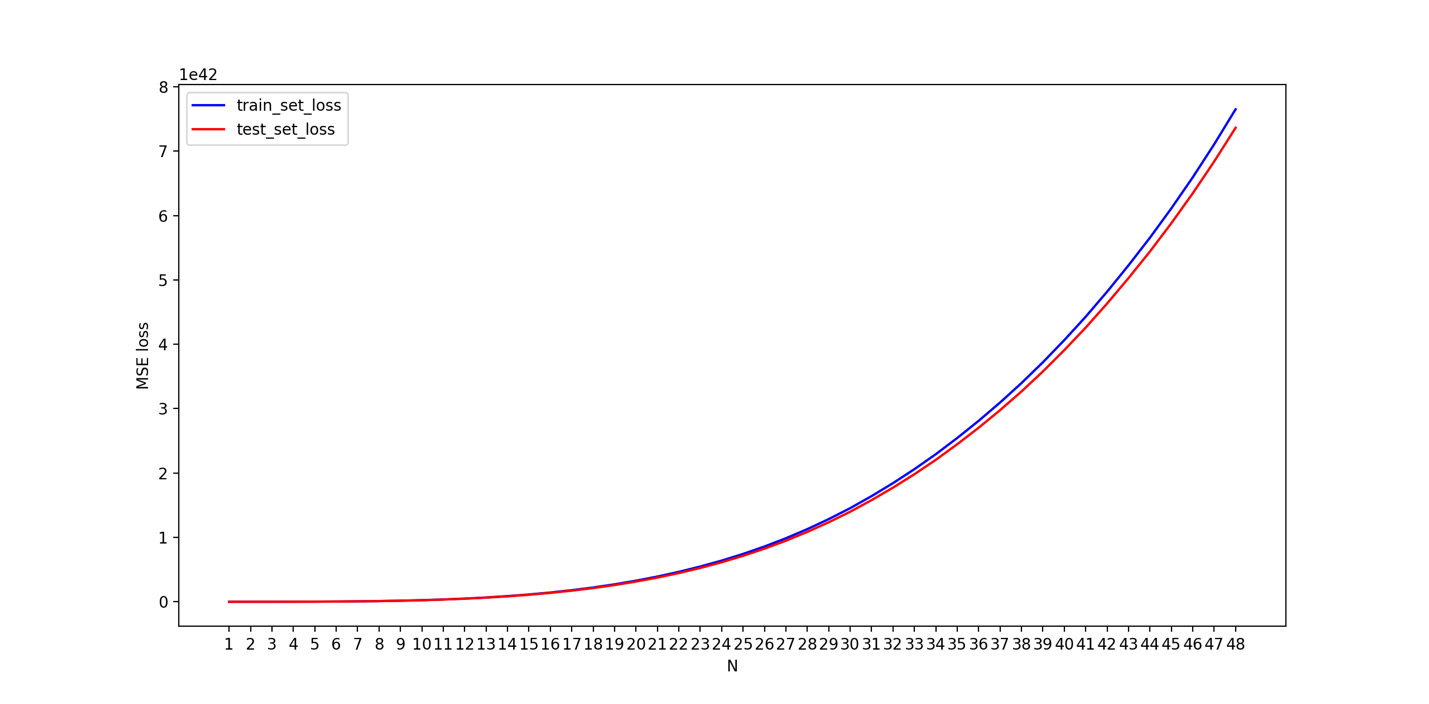
**Q2**

Plotting training error vs. testing error 

With training loss decreasing at a stable rate, yet testing loss increases perhaps due to overfitting.

**Q3**

Ex. Why do we need matrix inversion in linear regression. Can you proceed without it?

A: Without inversion the loss would never converge. 

At each data point, using the coefficients W results in some error of prediction, so we have n prediction errors. e(W) = y – W\*x. And the MSE function is MSE(W) = , and since the matrix form e will result in 🡪 MSE(W) = e = after substituting e with . Since we want to find the point where the gradient of MSE = 0 , = 0 , and the weights should be set to . Where the inversion sign shows.