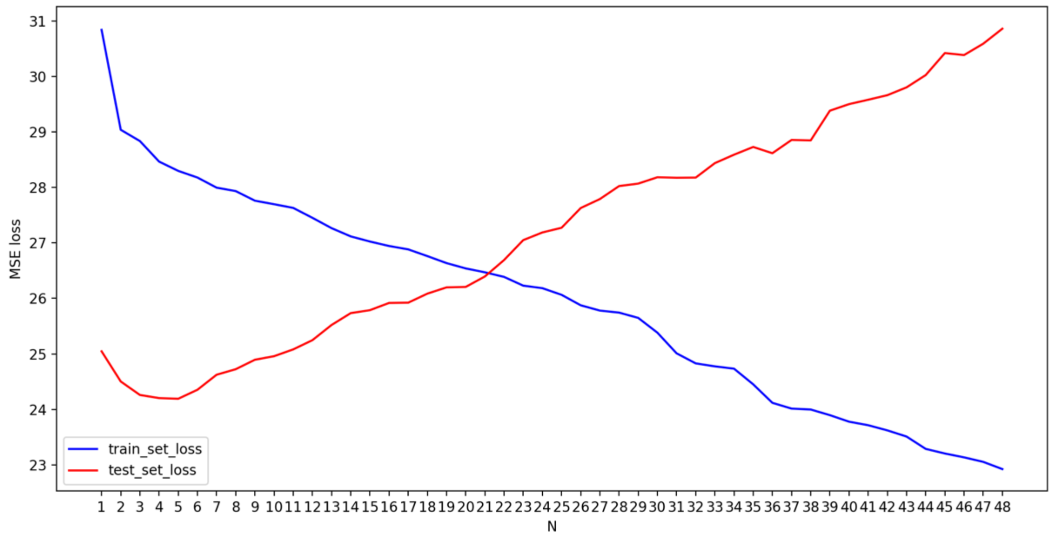
Q2

Plotting training error vs. testing error

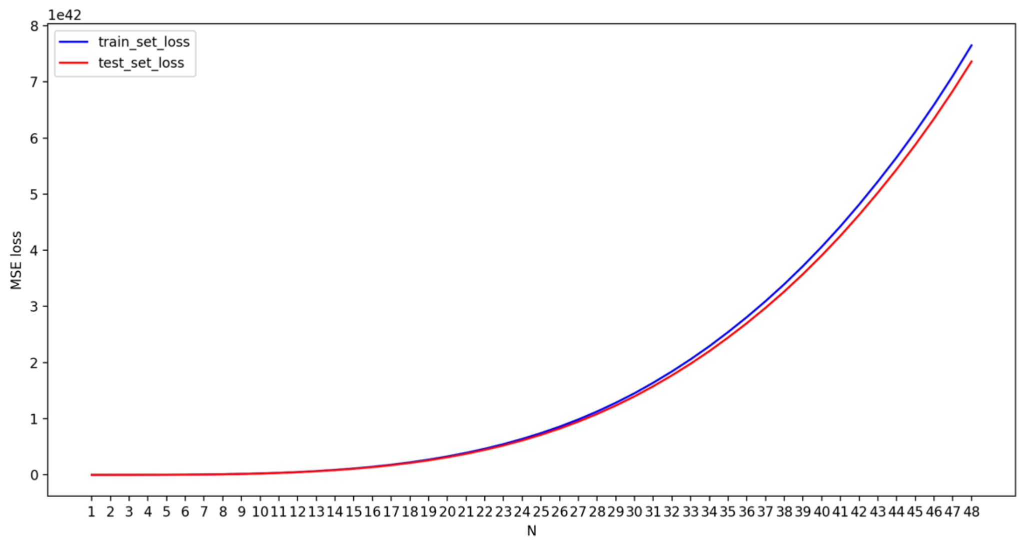


With training loss decreasing at a stable rate, yet testing loss increases perhaps due to overfitting.

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) = $\frac{1}{n}\sum e^2$, and since the matrix form e^Te will result in $\sum e^2$ → MSE(W) = $\frac{1}{n}e^T$e = $\frac{1}{n}(y^Ty - 2W^TX^Ty + W^TX^TXW)$ after substituting e with $y - Wx$. Since we want to find the point where the gradient of MSE = 0 , $\frac{1}{n}(\nabla y^Ty - 2\nabla W^TX^Ty + \nabla W^TX^TXW)$ = 0 , and the weights should be set to $(x^Tx)^{-1}x^Ty$. Where the inversion sign shows.