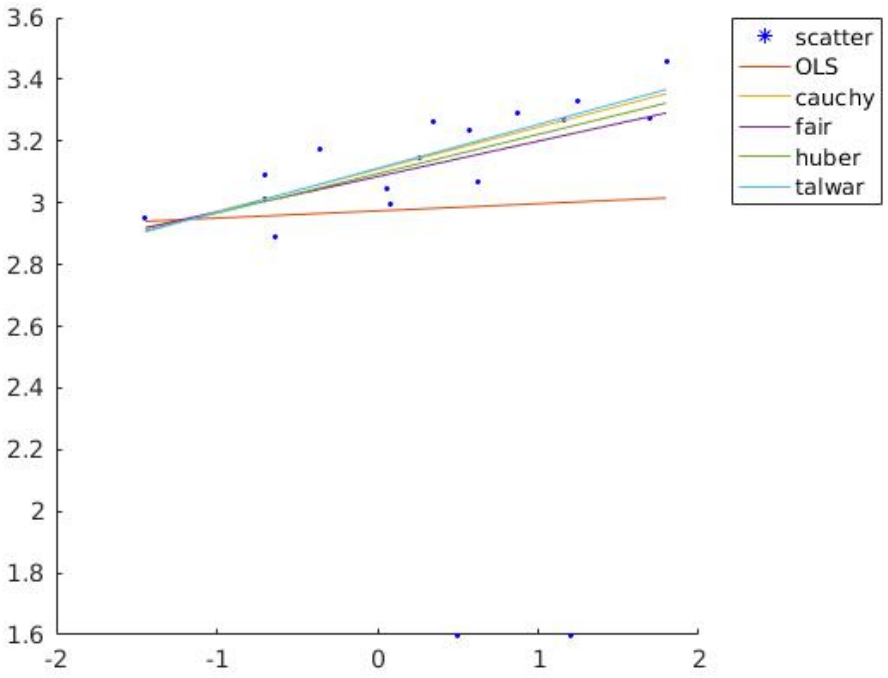
Problem 7.1
a): The OLS method minimizes the sum of squared residuals.
If X is the extension of input data, Then:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} = \left(\sum \mathbf{x}_i \mathbf{x}_i^\top \right)^{-1} \left(\sum \mathbf{x}_i y_i \right).$$

It would return a unique solution.
B = 2.9738, w = 0.02336
MSE_OLS = 0.2588 MAE_OLS = 0.3175

b)
MSE_OLS = 0.2588 MAE_OLS = 0.3175
MSE_cau = 0.2995 MAE_cau = 0.2427
MSE_fair = 0.2861 MAE_fair = 0.2468
MSE_hub = 0.2922 MAE_hub = 0.2451
MSE_tal = 0.3024 MAE_tal = 0.2435

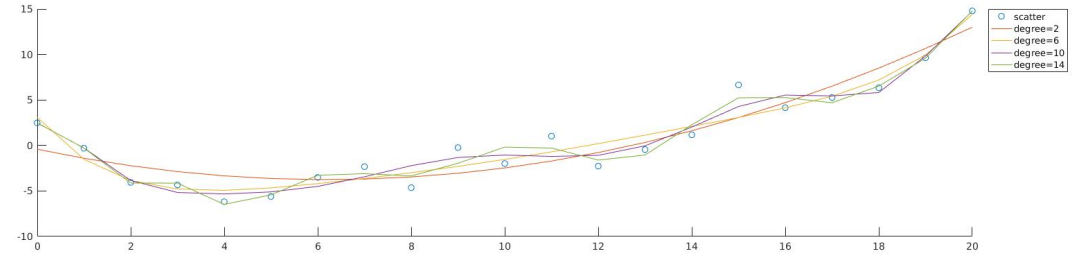
b_huber = 3.094 w_huber = 0.127

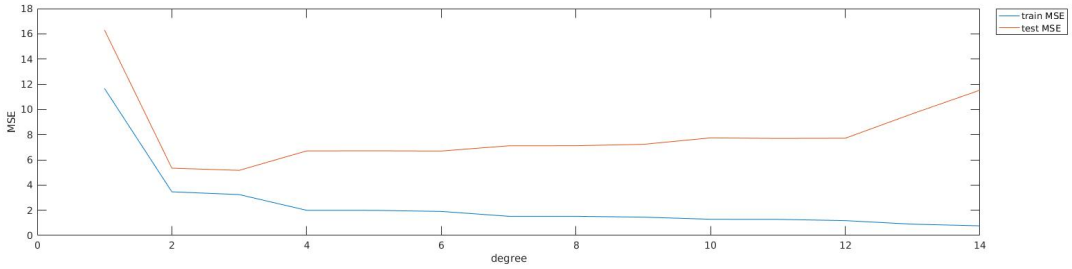


The outlier impact the OLS. Although OLS has a smallest MSE, it's MAE is larger, and it can not fit the other data point as well as robust regression

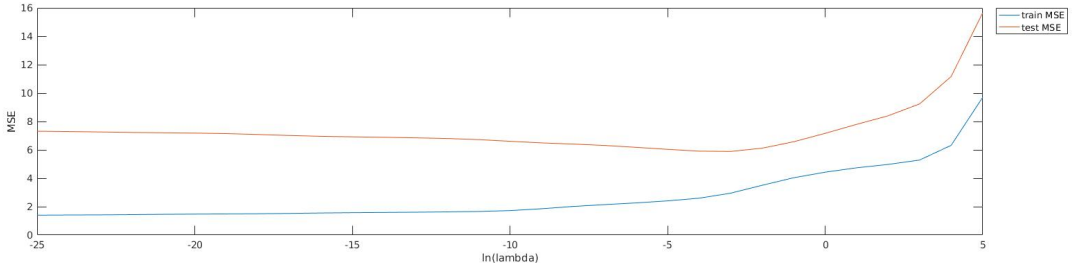
Problem 7.2

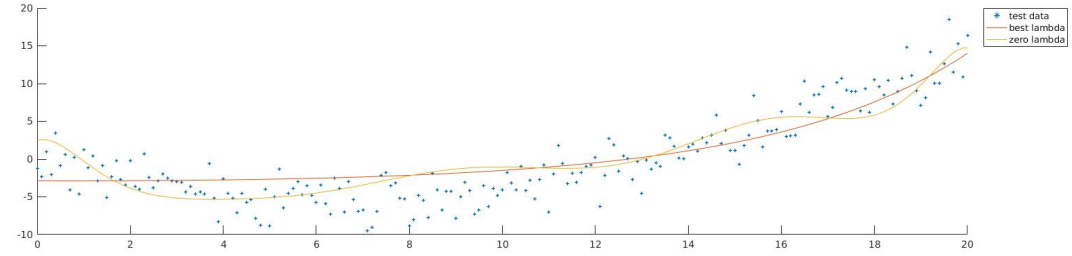
a) As the degree of the model growing, the curve become serpentine. The traing MSE decrease but the testing MSE increase. It means that the overfitting problem arise if the model is too complicated.



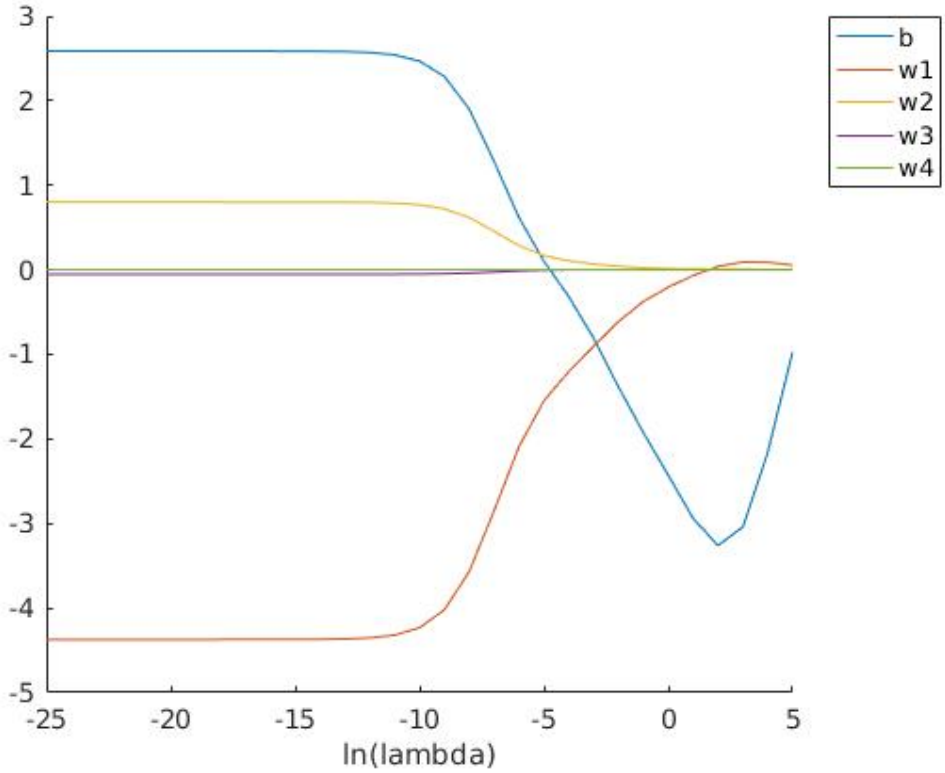


b) the testing MSE get its minimun at ln(lambda) = -3. With regulization, the cruve is smoother.



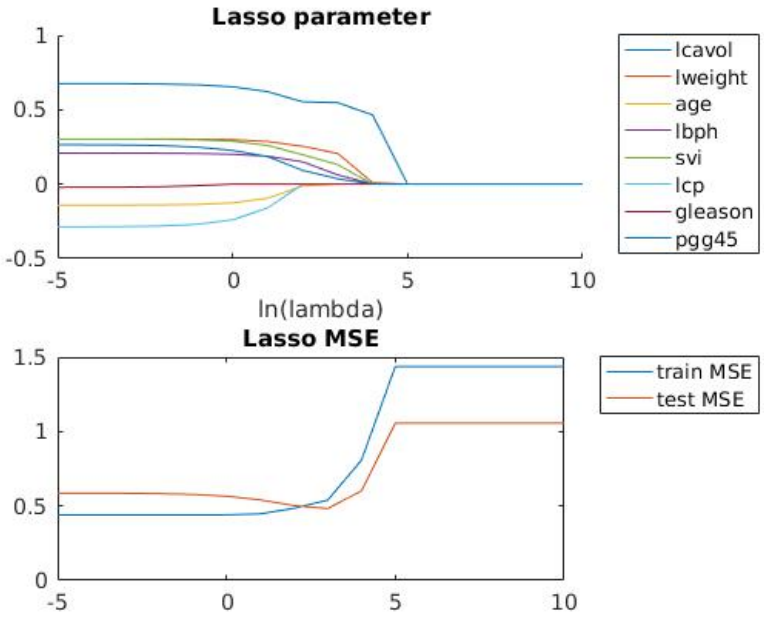


c) As the regulization parameter growing, all the weight parameter converge to zero.



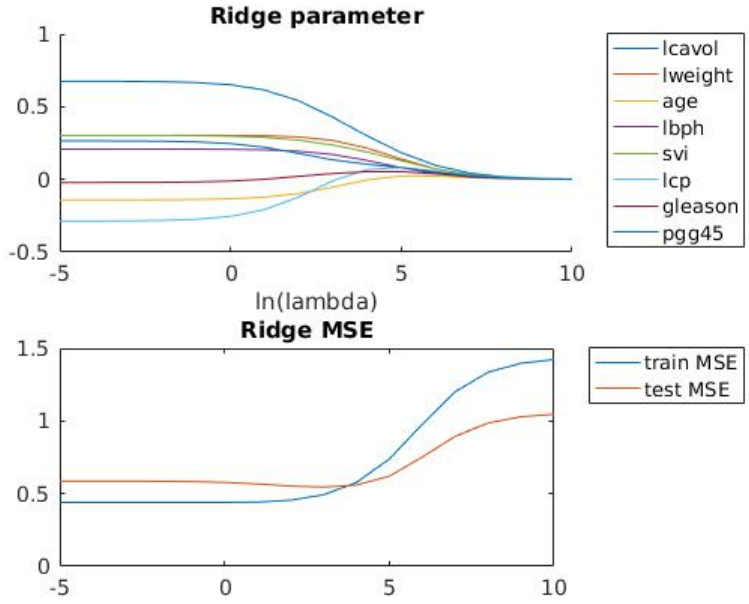
Problem 7.3

a):



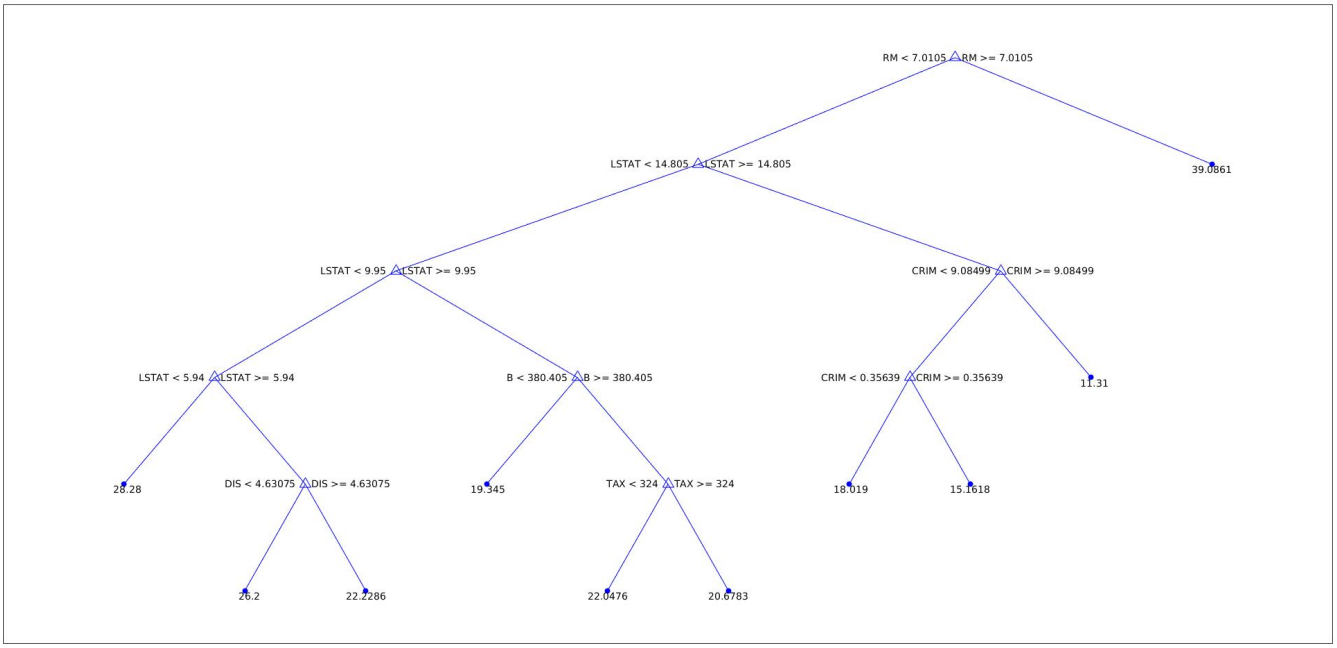
b) When Lasso lambda get large, all the parameter converge into 0. The two most sinificent coeffieient are lcavol and lweight. This can be use to reduce the parameter we use in predict model to avoid the overfitting problem.

c):



d): Ridge regression will also converge coeffient into 0, However, they converge smoothly and it require a lager lambda to get absolute 0 for Rigid regression

Problem 7.4
a) The regression tree is:



b): the predict is 22.0476.

c): When the observation per leaf is small, the training MAE is small but the testing MAE is large. It shows that overfitting happen when we set no limitation on small min observation on leaves. When the min leaf grow very large, the performance of the model get worse. The best minleaf parameter is around 15.