Alexa Summers

Santhoshini Sree Bolisetty

Gireesh Kumar Muppalla

Homework Chapter 6 & 7

1a) Best subset selection. This is because the other two rely on path dependency of which predictors to pick first.

1b) Undefined. The best subset could be the answer because it considers more models than the other methods, but the other methods might have better luck in picking a well-fit model.

1c) i. True

ii. True

iii. False

iv. False

v. False

2a) iii—less flexible and better predictions (less variance, more bias)

2b) iii—less flexible and better predictions (less variance, more bias)

2c) ii—more flexible, less bias, more variance

3a) iii—steadily increase. As we add together the values, the bias will continue to go up. Once we work with the testing data, the error will go down.

3b) ii—decrease initially, and then eventually start increasing in a U shape. This is the lasso penalty graph. As we increase s, the RSS decreases, and eventually as βs approaches their max, the they will begin overfitting to the training data, and test RSS increases

3c )iii—steadily increase

3d) iii—steadily increase.

3e) v—remains constant. Irreducible error is model independent.

4a) iv—steadily decrease. If lambda increases,

4b) decrease initially, and then eventually start increasing in a U shape.

4c) steadily decrease.

4d) steadily increase.

4e) Remain constant.

5.

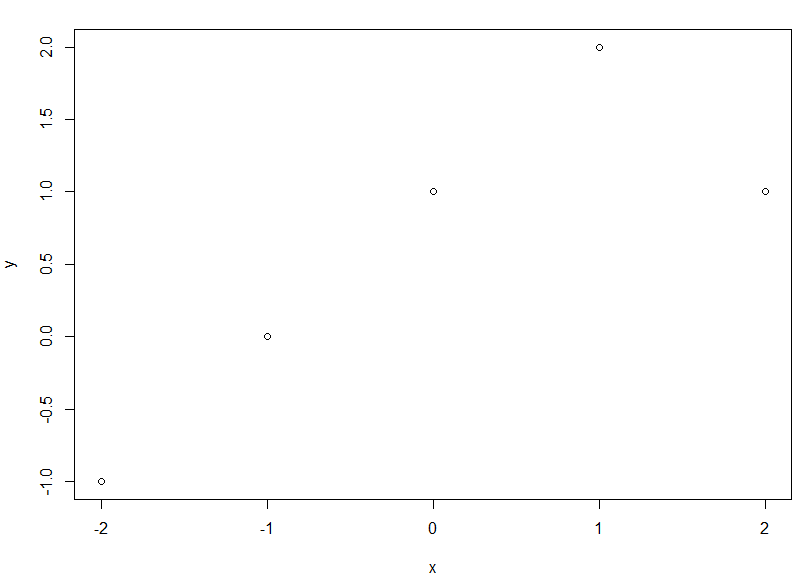
6a) g = 0 || g(0) -> 0

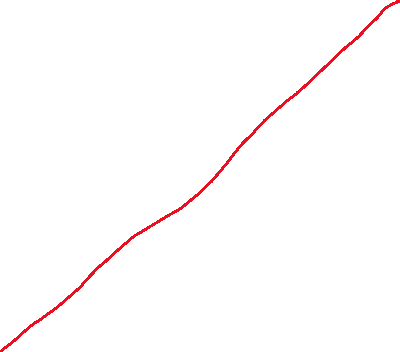
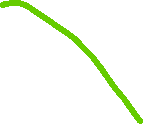
6b) g = c || g(1) -> 0

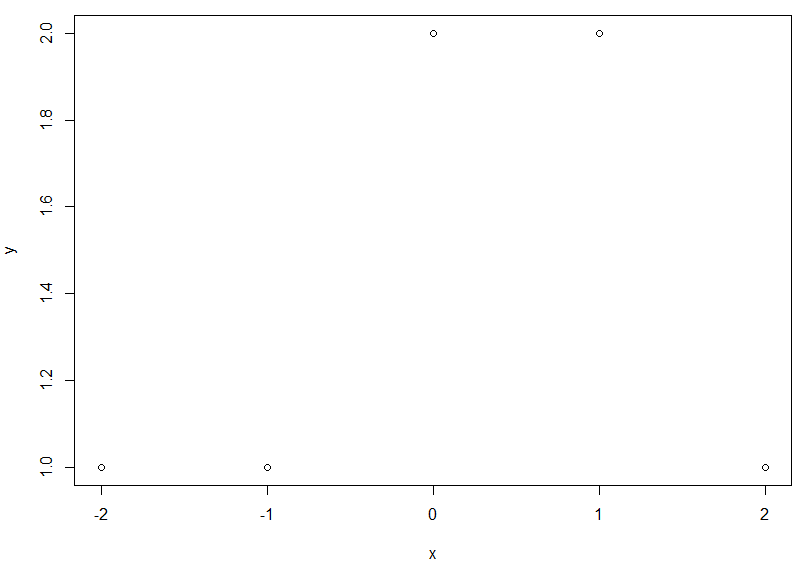
6c) g = cx + d || g(2) -> 0

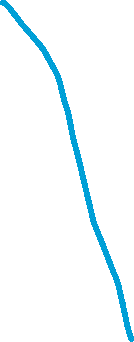
6d) g = cx2 +dx+ e || g(3) -> 0

6e) Penalty term does not do anything, so g is the interpolating spline.

7) Curve is linear between -2 and 1 (y = 1 + x). Curve is quadratic between 1 and 2 (y=1+x-2(x-1)2)



8) Curve is constant between -2 and 0 (y = 1), and between 0 and 1 (y =2). Curve is linear between 1 and 2 (y = 3 – x).



9a) g2 would probably have a smaller training RSS because it will be a higher order polynomial due to the order of the penalty term.

9b) g1 would probably have the smaller test RSS because g2 might overfit since it’s more flexible.

9c) If λ=0, then g1=g2, so the training and test RSS will be the same.