2a). The lasso, relative to least squares, is: iii. Less flexible and hence will give improved prediction accuracy when its increase in bias is less than its decrease in variance. This is because the lasso can yield a reduction in variance when compared to least squares in exchange for a small increase in bias, consistently generating more accurate predictions, also making it easier to interpret, making iii the correct choice.

2b). The ridge regression, relative to least squares, is: iii. Less flexible and hence will give improved prediction accuracy when its increase in bias is less than its decrease in variance. This is because, similarly to lasso, the ridge regression can yield a reduction in variance, when compared to least squares, in exchange for a small increase in bias. The relationship between λ and variance and bias is important: when it increases, the flexibility of the ridge regression decreases which causes decreased variance, but increased bias, making iii the correct choice again.

2c). Non-linear methods, relative to least squares, is: ii. More flexible and hence will give improved prediction accuracy when its increase in variance is less than its decrease in bias. Contrasting to ridge and lasso methods, non-linear methods work in the opposite way, giving increased prediction accuracy when a decrease in bias gives way to an increase in variance, making ii the correct choice.

3a). As we increase s from 0, the training RSS will: iv. Steadily decrease. As we begin to increase s from 0, all beta’s will increase from 0 to their least square estimate values. The training RSS for beta’s at 0 will be the maximum and trend downward to the original least squares RSS; therefore, iv is the correct choice.

3b). As we increase s from 0, the test RSS will: ii. Decrease initially, and then eventually start increasing in a U shape. When s=0 and all beta’s are 0, the model is extremely simple and because of that, has a high test RSS. Beginning to increase s, beta s will begin to assume non-zero values and the model begins to fit better, so test RSS originally decreases. Eventually, beta s will approach their OLS values, and as they begin to overfit the training data, test RSS will begin to increase again, forming a U shape and making ii the correct choice.

3c). As we increase s from 0, the variance will: iii. Steadily Increase. When s=0, the model basically predicts a constant and has almost no variance, but as we increase s, the model includes more Betas, and their values will begin to increase. As the values of Betas become increasingly more dependent on training data, the variance will steadily increase, making iii the correct choice.

3d). As we increase s from 0, the (squared) bias will: iv. Steadily Decrease. As we stated in the previous example, when s=0, the model basically predicts a constant, so the prediction is far from the actual value, and (squared) bias is high. As we increase s from 0 though, more Beta’s become non-zero, and the model continues to fit the training data better, thus making bias steadily decrease, and proving iv to be the correct choice.

3e). As we increase s from 0, the irreducible error will: v. Remain Constant. Irreducible error is model dependent and therefore increasing s from 0 will not change it, making it remain constant and proving v to be the best choice.

10).