Theoretical Part

1.

(1) For Ridge Regression:

$$||y - X\beta||^2 + \lambda \sum_{k=1}^{p} {\beta_k}^2$$

When λ increases from 0:

(i) training RSS: steadily increase. Because as λ gets larger, the coefficients become shrunken towards zero. We will have less variables to fit the data. Therefore, the training RSS will steadily increase.

(ii): test RSS: Increase initially and then eventually start decreasing in an inverted U shape. At first, λ equals to 0. In this situation, we consider as using least square methods to fit the data using all the variables. Then, the RSS will increase. As λ is increasing, the number of β which will be zero will also increase. Thus, the over-fitting problem will be solved. Therefore, the RSS will decrease.

(iii): Variance: Steadily decrease. When λ equals to 0, we use least square method to fit the data, and we include all the variables. Therefore, the variance may be large. As λ increases, the coefficients will become shrunken and decrease to 0, and we have less variables, When λ is infinity, we will have no variables and the variance becomes approximately 0.

(iv): (squared) bias: Steadily increase. According to the formula,

Test Error = Bias2 + Variance

When λ is 0, we have all the variables including true model's variables. In this situation, the bias2 is the smallest. As λ starts to increase, β will start to decrease, therefore, we may have less variables to fit the data. When λ becomes infinity, we have no variables and bias2 will be the largest.

(2) for Lasso Regression

$$||y - X\beta||^2 + \lambda \sum_{j} |\beta_j|$$

Equivalently, find β that minimizes

$$||y - X\beta||^2$$

Subject to the constraint that

$$\sum_{j=1}^{p} |\beta_j| \le s$$

(i): training RSS: Steadily decease. If s increases from 0, when s is extremely large, the results will be estimates of using least square method. Therefore, When s equals zero, there will be no betas, and training RSS is the biggest. When s starts to increase, the training RSS will decrease, and when s is close to infinity, it will reach to the OLS RSS

(ii): test RSS: Decrease initially, and then eventually start decreasing in an inverted U shape. When s=0, there will be no betas, and the test RSS is the biggest. As s starts to increase, the number of non-zero betas will start to decrease, and test RSS will begin to decrease. While s continues to increase, the over-fitting problems will appear, and RSS will begin to increase.

(iii): Variance: When s=0, the model has no betas and is the simplest, therefore the variance is the smallest. While s starts to increase, the number of non-zero betas will decrease, and the variance will start to increase.

(iv): (squared) bias: When s=0, the model has no betas, and is not a true model, therefore the bis is the biggest. While s starts to increase, the model is more close to true model, the bias will start to decrease.

2. (25 pt) This problem illustrates the estimator property in the shrinkage methods. Let Y be a single observation. Consider Y regressed on an intercept

$$Y = 1 \cdot \beta + \epsilon$$

(a) Using the formulation as shown in class, write down the optimization problem of general linear model, ridge regression and LASSO in estimating β respectively.

General linear model:

$$\min_{\beta} = (Y - \beta)^2$$

Ridge Regression:

$$\min_{\beta} = (Y - \beta)^2 + \lambda \beta^2$$

Lasso Regression:

$$\min_{\beta} = (Y - \beta)^2 + \lambda |\beta|$$

(b) For fixed tuning parameter λ, solve for β (general linear model), β_λ (ridge regression) and β_λ

(LASSO) respectively.