3. Suppose we estimate the regression coefficients in a linear regression model by minimizing

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 \quad \text{subject to} \quad \sum_{j=1}^{p}|\beta_j| \le s$$

for a particular value of s. For parts (a) through (e), indicate which of i. through v. is correct. Justify your answer.

(a) As we increase s from 0, the training RSS will:

    i. Increase initially, and then eventually start decreasing in an inverted U shape.

    ii. Decrease initially, and then eventually start increasing in a U shape.

    iii. Steadily increase.

    iv. Steadily decrease.

    v. Remain constant.

(b) Repeat (a) for test RSS.

(c) Repeat (a) for variance.

(d) Repeat (a) for (squared) bias.

(e) Repeat (a) for the irreducible error.

Ans:

(a) iv (4)

(b) ii (2)

(c) iii (3)

(d) iv (4)

(e) v (5)

Reason

(a) iv : When s increases, the model's constraints decrease, allowing it more freedom to fit data. The RSS will steadily decrease.

(b) ii : When s increases, the complexity of model rises, test error will decrease when model gets to fit the data better. But it will cause **overfitting** when model becomes more complex, then test RSS increases.

(c) iii : As s increases, the complexity of model rises. and freedom in parameter estimation increases, leading to **higher variation**.

(d) iv : Increasing s reduces bias as model becomes more flexible and better able to fit the data.

(e) v : Because it depends on the random noise in data not the complexity, it remains constant.