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

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

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.

When s starts to increase from 0, more coefficients are allowed to become non-zero, enhancing the model's ability to fit the data, leading the training RSS (error) to steadily decrease. **(iv)**

(b) Repeat (a) for test RSS.

As s increases, the model becomes more complex, reducing training error.
However, on the test data, the error may initially decrease (better generalization), but as model complexity continues to grow, overfitting may occur, causing the test RSS to exhibit a **U-shaped increase**. **(ii)**.

(c) Repeat (a) for variance.

As s increases, the model gains more degrees of freedom, making it more sensitive to variations in the training data, and thus the **variance will steadily increase**. **(iii)**.

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

When s increases, the model becomes better at fitting the data, and bias is expected to **steadily decrease** as the model complexity grows. **(iv)**.

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

The irreducible error is due to noise and other unexplained variance in the data and is not affected by changes in sss. Hence, it will **remain constant**. The correct answer is **(v)**.