

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  $s$  increases:

- Increasing  $s$  allows for larger values of the coefficients  $\beta_j$ , allowing the model to fit the training data more closely. This means that the training RSS will **decrease** as  $s$  increases from 0.

b.)

As  $s$  increases from 0, the **test RSS will initially decrease**, but after a certain point (when the model starts overfitting), the **test RSS will start increasing** again, creating a **U-shaped curve**.

**Answer**

The correct answer is: ii. **Decrease initially, and then eventually start increasing in a U shape.**

c.)

As  $s$  increases, the **variance of the model's predictions will steadily increase** because the model becomes more complex and flexible.

**Answer**

The correct answer is: iii. **Steadily increase.**

d.)

As  $s$  increases, the **squared bias of the model's predictions will steadily decrease** because the model becomes more flexible and better able to fit the data.

**Answer**

The correct answer is: iv. **Steadily decrease.**

e.)

The irreducible error **remains constant** regardless of changes to  $s$ .

### **Answer**

The correct answer is: v. **Remain constant.**