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) For training RSS：

Answer: **iv. Steadily decrease.**

Explanation:

When s=0, all coefficients βj are forced to be zero, so the model predicts only the intercept β0 (model can't fit the data well without features), yielding high training RSS.

As s increases, more flexibility is allowed in the coefficients, reducing the training RSS because the model can fit the data better.

(b) For test RSS：

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

Explanation:

With very low s, the model is too constrained, resulting in underfitting and high test RSS. As s increases, the model can fit the data better, reducing test RSS. However, after a certain point, further increasing s leads to overfitting, causing test RSS to rise again.

(c) For variance：

Answer: **iv. Steadily increase.**

A model's variance is related to it's sensitivity to dataset. When a model's variance is high, it means that the model is sensitive to dataset.

Explanation:

With a small s, the coefficients (βj) are heavily constrained, leading to low variance. As s increases, the model has more flexibility, causing higher variance in the estimated coefficients (may lead to overfitting).

(d) For (squared) bias：

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

Explanation:

At very low s, the model is **overly simplistic** (highly biased). Increasing s reduces bias as the model becomes more flexible. However, too large s may cause the model to **capture noise**, which could indirectly affect the model's ability to generalize, potentially increasing bias slightly due to overfitting.

(e) For irreducible error：

Answer: **v. Remain constant.**

Explanation:

The irreducible error is due to **inherent noise in the data** and **cannot be affected by the model** or the choice of s. Therefore, it remains constant regardless of s.