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

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2$$

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

(a) As we increase λ 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.

As λ increases, the penalty term will grows, making the model smoother and shrinking the regression coefficients, which reduces the model's fitting ability and causes training RSS to steadily increase. (iii)

(b) Repeat (a) for test RSS.

As λ increases, the model becomes simpler, reducing overfitting, so the test RSS may initially decrease. However, if λ becomes too large, the model will underfits, causing test RSS to increase in a U shape. **(ii)**

(c) Repeat (a) for variance.

Explanation: As λ increases, the model becomes smoother and simpler, reducing its sensitivity to data fluctuations, so **variance steadily decreases**. **(iv)**.

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

As λ increases, the model complexity decreases, causing **bias to steadily increase** because the model's ability to fit the data is reduced. **(iii)**.

(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 λ\lambdaλ. Hence, it will **remain constant**. The correct answer is **(v)**.