

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:
- Increase initially, and then eventually start decreasing in an inverted U shape.
 - Decrease initially, and then eventually start increasing in a U shape.
 - Steadily increase.
 - Steadily decrease.
 - 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.

(a) Training RSS (Residual Sum of Squares):

- As λ increases, regularization shrinks β_j , reducing the model's flexibility.
- Answer: (iii) Steadily increase.**
Higher λ leads to simpler models that may underfit, increasing training error.

(b) Test RSS:

- Test RSS typically decreases initially (reduced overfitting) but increases later (underfitting) as λ grows.
- Answer: (ii) Decrease initially, and then eventually start increasing in a U shape.**

(c) Variance:

- Variance decreases as λ grows because the model becomes less sensitive to small changes in the data.
- **Answer: (iv) Steadily decrease**

(d) (Squared) Bias:

- Bias increases with larger λ since the model becomes too simple and underfits.
- **Answer: (iv) Steadily increase.**

(e) Irreducible Error:

- Irreducible error is inherent in the data and does not depend on λ .
- **Answer: (v) Remain constant**