

**Data Mining
Concept Quiz
Oct 11**

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.
- (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)

Will steadily increase(iii).

An increase in lamda will cause increasing restrictions to beta, causing the model to become less flexible. This will lead to increased RSS.

b)

It will decrease initially then will slowly transform into a U shape

(ii).

While the model is become less and less flexible, the RSS will decrease first and start increasing when overfitting occurs.

c)

It will steadily decrease(iv), as model flexibility decrease variance will decrease.

d)

It will steadily increase(iii). Bias will increase with as model flexibility decrease.

e)

It will remain constant(v).

Irreducible error is independent of model parameter and therefore independent of lamda.