Model Order Selection. In-class Exercise 3

EL-GY 6143 Intro Machine Learning. Prof. Sundeep Rangan

Question

A teacher tries to model the score of a child on a test as follows:

$$Test\ score \approx \beta_0 + \beta_1[Hours\ studied] + \beta_2[Hours\ sleep\ before\ test]$$

Below are three possible shortcomings of the model. For each shortcoming, indicate whether it would lead to (i) High irreducible error, (ii) High bias error or (iii) High variance error

(a) The relation between test score and hours studied is not linear. After some amount of studying, the students test score will no longer improve.
(b) The model should also include the age of the child.
(c) The model was fit with data from one class with only 15 children.

Solution

(a) **High bias error** since the true relation between test score and hours studied is more complex than the simple linear relationship.
(b) **High irreducible error**. There is variation in the test score not explained by the two predictors being used.
(c) **High variance error**. High variance occurs when there are too few samples to fit the model.