

I have the cost function

$$C(X, \beta) = E[(Y - \hat{Y})^2]$$

Will show that it can be rewritten

as $E[(Y - \hat{Y})^2] = \text{Bias}[\hat{Y}] + \text{Var}[\hat{Y}] + \sigma^2$

where

$$\text{Bias}[\hat{Y}] = E[(Y - E[\hat{Y}])^2]$$

and

$$\text{Var}[\hat{Y}] = E[(\hat{Y} - E[\hat{Y}])^2] = \frac{1}{n} \sum_i (Y_i - E[\hat{Y}])^2$$

$$E[(Y - \hat{Y})^2] = E[Y^2] - 2E[Y\hat{Y}] + E[\hat{Y}^2]$$

where $Y = f + \epsilon$

$$\begin{aligned} E[Y^2] &= E[(f + \epsilon)^2] \\ &= E[f^2 + 2f\epsilon + \epsilon^2] \\ &= E[f^2] + \sigma^2 = f^2 + \sigma^2 \end{aligned}$$

$$\begin{aligned} E[Y\hat{Y}] &= E[(f + \epsilon)\hat{Y}] \\ &= E[f\hat{Y} + \epsilon\hat{Y}] \\ &= f E[\hat{Y}] \end{aligned}$$

$$E[\hat{Y}^2] = \text{Var}[\hat{Y}] + E[\hat{Y}]^2$$

So now,

$$\begin{aligned} E[(Y - \hat{Y})^2] &= f^2 - 2fE[\hat{Y}] + E[\hat{Y}]^2 \\ &\quad + \text{Var}[\hat{Y}] + \sigma^2 \end{aligned}$$

where

$$\begin{aligned} &f^2 - 2fE[\hat{Y}] + E[\hat{Y}]^2 \\ &= E[(f - E[\hat{Y}])^2] \\ &= E[(Y - E[\hat{Y}])^2] = \text{Bias}[\hat{Y}] \end{aligned}$$

$$\Rightarrow E[(Y - \hat{Y})^2] = \text{Bias}[\hat{Y}] + \text{Var}[\hat{Y}] + \sigma^2$$

A high bias indicates that

the population parameter differs greatly from the corresponding estimators expected values. It often introduces systematic errors, and underfitting due to high bias does not allow the model to consider finer details in the data (like curvature).

In other words, how far off the mark the model is as a whole.

A high variance indicates that

an estimator tends to vary greatly from its expected value (i.e. its not very tight). If a model is overly sensitive to the training data, it may attempt to fit the noise, leading to overfitting.

In other words, how flexible the model is for a given data set.