→ whenever we discuss model prediction we have to understand prediction errors (bias and variance)

→ Gaining a proper understanding of these errors helps us to build accurate models but also avoid the mistakes as overfitting and underfitting.

→ Bias ?

Bias is the difference between the average prediction of our model and the correct value which we are trying to predict. Models with high bias pays very little attention to the training data and oversimplifies the model.

Variance?

Variance is a value which tells us the spread of our data. Models with high variance pays a lot of attention to the training data and do not generalize on the data which it hasn't seen before.

Mathematically :-

Lets assume that the relationship betw een the predicted y and x is defined a

$$y = f(x) + e$$

where $c \rightarrow$ error
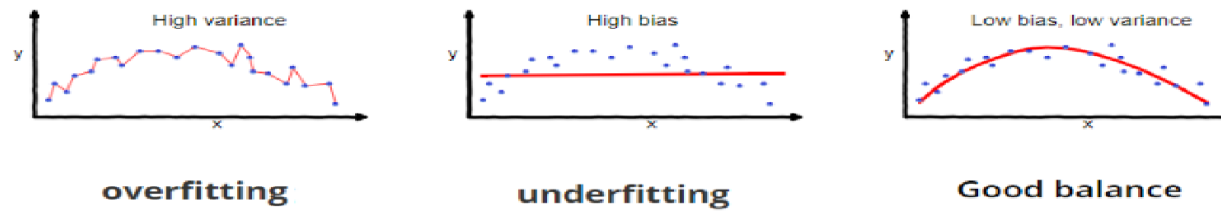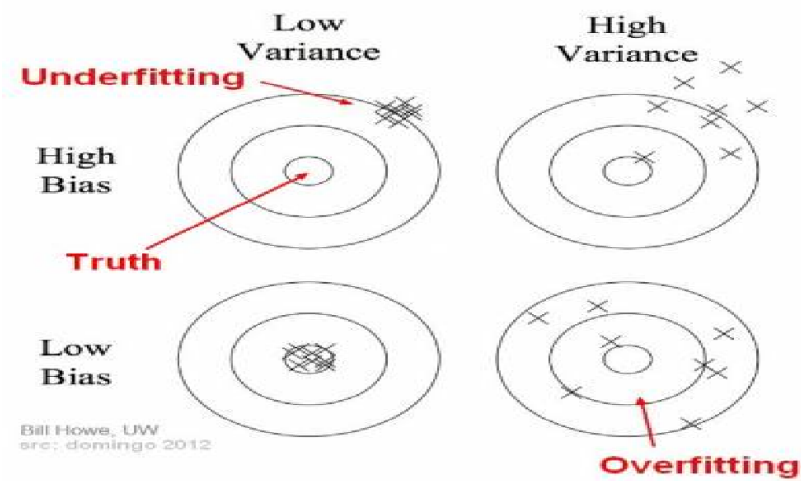
$$f(x) \rightarrow \text{function of } x$$

After making the model $\hat{f}(x)$ of $f(x)$ using regression or other technique we can calculate the error as.

$$Err(x) = E\left[(y - \hat{f}(x))^2\right]$$

The $Err(x)$ is further decomposed into:

$$Err(x) = \left(E[\hat{f}(x)] - f(x)\right)^2 + E\left[\left(\hat{f}(x) - E[\hat{f}(x)]\right)^2\right] + \sigma_e^2$$

$$Err(x) = Bias^2 + Varriance + Irreducible\ Error$$

Low Variance — High Variance

Underfitting

High Bias

Truth

Low Bias

Bill Howe, UW
src: domingo 2012

Overfitting

High variance

High bias

Low bias, low variance

overfitting   **underfitting**   **Good balance**

Total Error = Bias^2 + Variance + Irreducible Error

Error

Optimal Balance

Total Error

Bias^2

Variance

Algo Complexity