



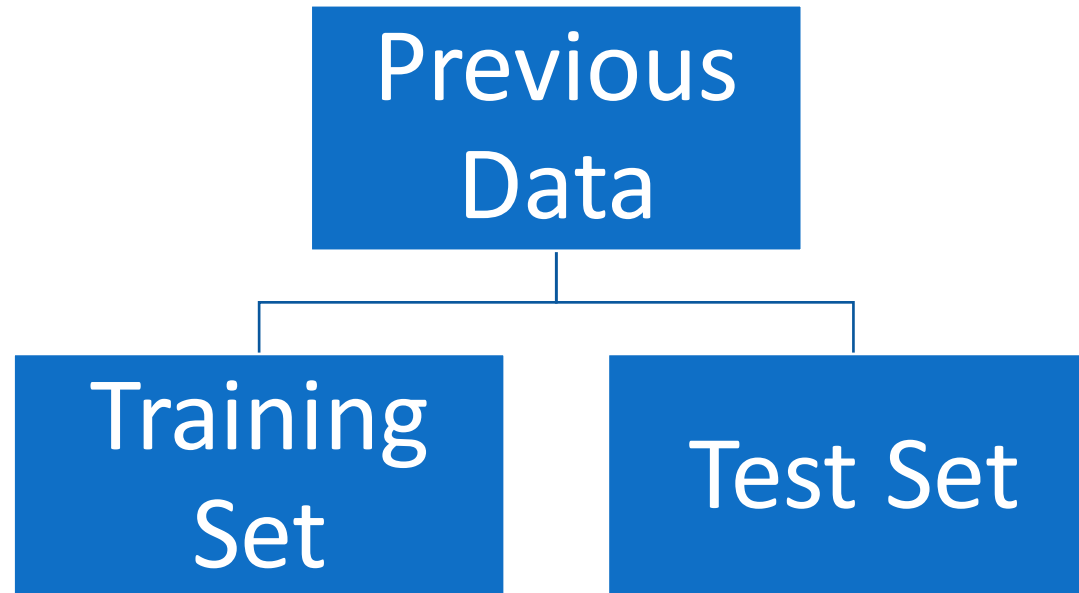
Start-Tech Academy

# Linear Regression

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2$$

- Training error – Performance of model on the previously **seen** data
- Test error – Performance of model on the **unseen** data

## Test-Train Split



# Linear Regression

## Test-Train Split

Training Set -  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$

Model is trained

$$y = f(x)$$

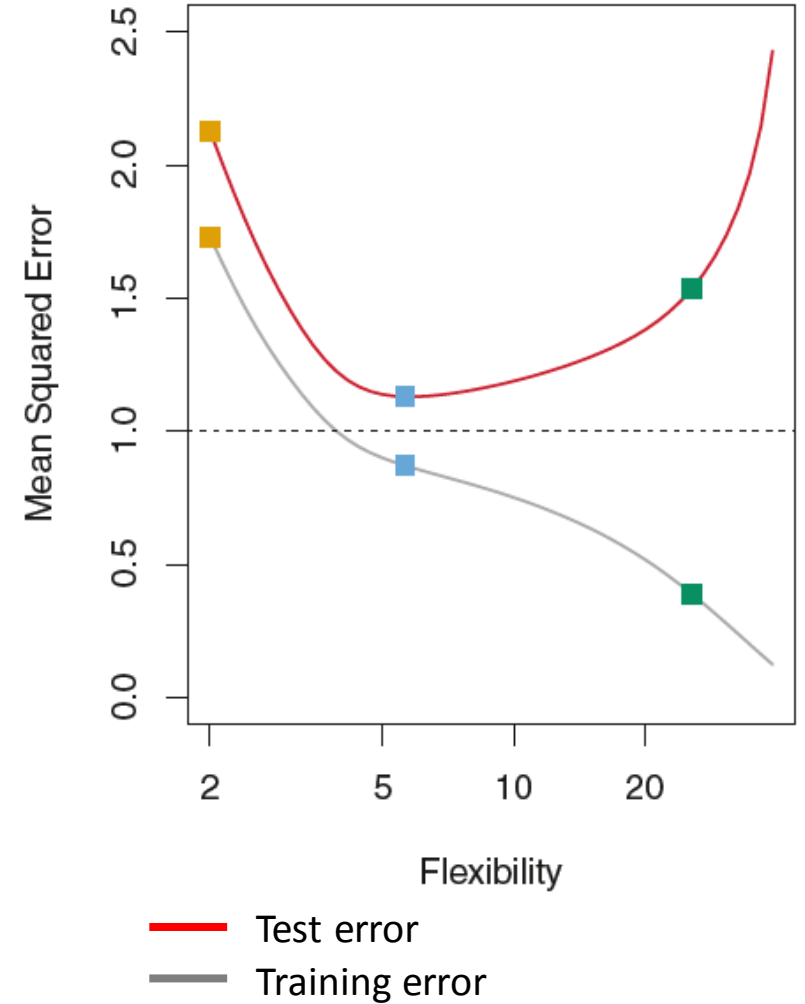
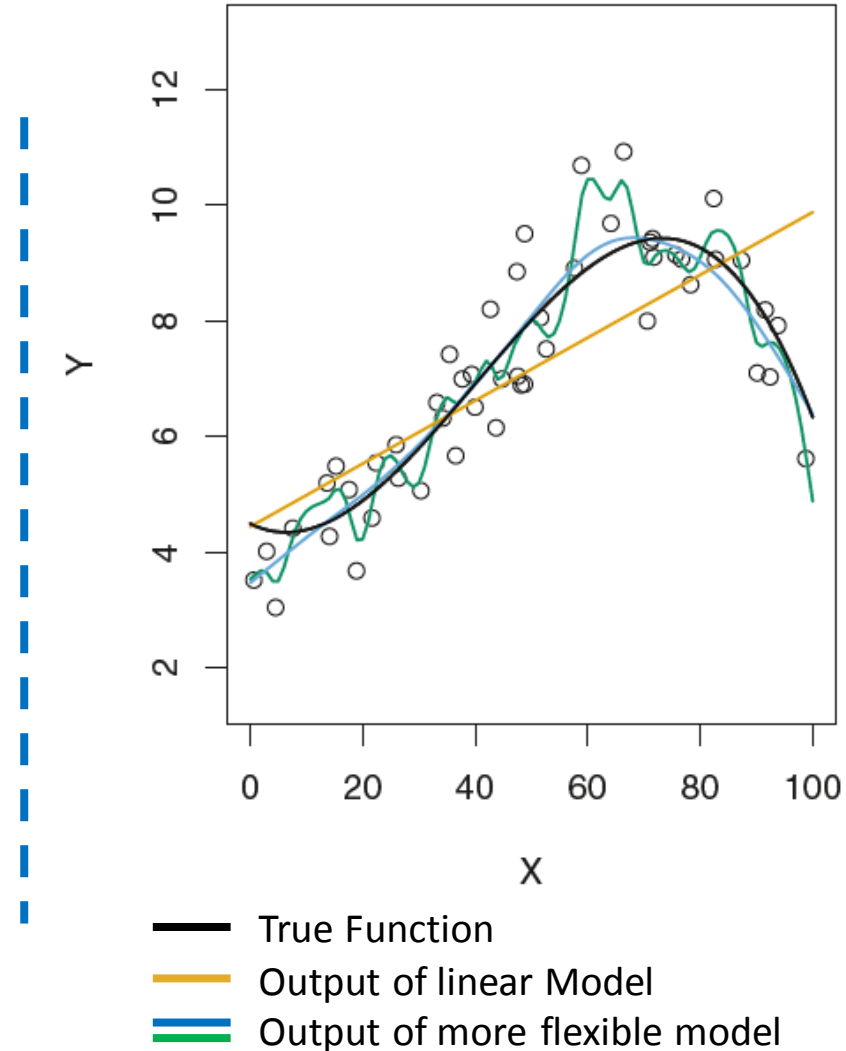
Test Set - Previously unseen data  $(x_0, y_0)$

Test MSE -  $\text{Ave}(\hat{f}(x_0) - y_0)^2$



# Other Linear Regression

## Test-Train Split



# Linear Regression

## Test-Train Split Techniques

### 1. Validation set approach

- Random division of data into two parts
- Usual split is 80:20 (Training : Test)
- When to use – In case of large number of observations

### 2. Leave one out cross validation

- Leaving one observation every time from training set

### 3. K-Fold validation

- Divide the data into k set
- We will keep one testing and K-1 for training

