Estimate $\hat{f}$ = Learn $\hat{f}$

$$Y = f(X) + \epsilon$$

the squared error for a given estimate $\hat{f}$ is

$$E(\text{actual} - \text{predicted})^2 = E(Y - \hat{Y})^2$$

which factors as

$$E[f(X) + \epsilon - \hat{f}(X)]^2$$

$$\underbrace{[f(X) - \hat{f}(X)^2]}_{\text{reducible}} + \underbrace{\text{Var}(\epsilon)}_{\text{irreducible}}$$

Training

training data set

$$\{(y_1, x_1,), \ldots, (y_n, x_n))\}$$

used to find function q that minimizes **Training MSE**

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

Testing

testing data sets (unseen)

$$(y_0, x_0)$$

used to compute **Test MSE**

$$E[y_0 - \hat{f}(x_0)^2]$$

often not so closely related