

M339 G: September 15th, 2025.

Assessing Model Accuracy.

Say, we have our "usual" regression model

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

X and ϵ are independent,
and $\epsilon \sim N(0, \sigma^2)$

Say, we fit our model to some training data:

$$Tr = \{ (x_i, y_i) : i = 1, \dots, N \}$$

Let \hat{f} be the fit of the model to our Tr .

$$MSE_{Tr} = \text{Ave}_{i \in \{1, \dots, N\}} (y_i - \hat{f}(x_i))^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{f}(x_i))^2$$

We propose to look @ other data

$$Te = \{ (x_j, y_j) : j = 1 \dots M \}$$

These are our testing data.

We calculate

$$MSE_{Te} := \text{Ave}_{j=1 \dots M} (y_j - \hat{f}(x_j))^2 = \frac{1}{M} \sum_{j=1}^M (y_j - \hat{f}(x_j))^2$$