

80

20

x_1

x_2

x_3

y

\hat{y}

Residual

$y - \hat{y}$

$(y - \hat{y})^2$

1

2

3

4

5

—

—

—

—

—

—

—

—

—

—

$n = 100$

MSE =

9 →

Split the data → Training + Testing
 80% 20%

| | x_1 | x_2 | x_3 | y | \hat{y} | $(y - \hat{y})^2$ |
|-----|-------------|-------|-------|-------------|-------------------|-------------------|
| 1 | X_{train} | | | y_{train} | \hat{y}_{train} | MSE_{train} |
| 2 | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| 80 | X_{test} | | | y_{test} | \hat{y}_{test} | MSE_{test} |
| 81 | | | | | | |
| 100 | | | | | | |

Performs extremely well (Minimize) $MSE \approx 0$

Performs badly $MSE \uparrow \rightarrow 50$

- 1 `model = LinearRegression()`
- 2 `model.fit(X_train, y_train)`
- 3 `yhat_train = model.predict(X_train)`
- 4 `yhat_test = model.predict(X_test)`

Correct Fit \rightarrow low training - 3 }
low testing error - 5 }

"UNSEEN" Data