SIS 1.

1. Exercise: Comparing MSE and RMSE

True values: $y_i = [500, 300, 800, 400, 6000]$

Predicted values: $\hat{y}_i = [450, 350, 780, 420, 910]$

Step 1: Compute the squared errors $(y_i - \hat{y}_i)^2$:

1. $(500-450)^2 = 50^2 = 2500$
2. $(300-350)^2 = (-50)^2 = 2500$
3. $(800-780)^2 = (20)^2 = 400$
4. $(400-420)^2 = (-20)^2 = 400$
5. $(6000-910)^2 = (5090)^2 = 25908100$

Step 2: Compute the MSE

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

$$MSE = \frac{2500 + 2500 + 400 + 400 + 25908100}{5} =$$

$= 5182780$

Answer: $MSE = 5,182,780$.

3. What does the result quantify?

The MSE measures the average squared difference between the actual and predicted values. A higher MSE indicates a larger discrepancy between the predictions and the true values. Since the errors are squared, larger errors have a much greater impact on the final value. In this case, the MSE is 5,182,780 which is extremely high due to the large error in the last data point. This suggests that the model's predictions are significantly off for at least one value.

3. What are the disadvantages of this metric?

1) Sensitive to large errors (outliers)

- Since MSE squares the errors, it gives more weight to large errors, meaning a single large mistake can dominate the metric. This can make the model look worse than it actually is.

2) Not in the same unit as the target variable

- The MSE value is in squared units of the target variable, making it harder to interpret directly. For example, if the target is in dollars, MSE in dollars, which is not intuitive.

3) Does not indicate error direction.

- MSE only considers the magnitude of errors but does not tell us if the model is overestimating or underestimating the values.

Because of these drawbacks, we often use Root Mean squared error instead.

4. $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} = \sqrt{MSE}$

$RMSE = \sqrt{5182780} \approx 2276,57$

Answer: $RMSE \approx 2276,57$.

5. Interpretation of Results

The RMSE of 2276,57 indicates that, on average, the model's predictions are off by around 2276,57 units in the same scale as the data.

6. Difference Between MSE and RMSE and Why RMSE is Preferred

- MSE is in squared units, making it harder to interpret.
- RMSE is in the same units as the data, making it easier to understand.
- RMSE is preferred because it's more interpretable and directly reflects the average error in real-world terms, while MSE can be dominated by large errors.

Exercise: Bias Variance decomposition of MSE

$MSE = E[(\hat{Y} - Y)^2] = (Bias(\hat{Y}))^2 + Var(\hat{Y})$

Step 1: Expand the squared error

$(\hat{Y} - Y) = (\hat{Y} - E[\hat{Y}]) + (E[\hat{Y}] - Y)$

Substitute this into the squared term:

$(\hat{Y} - Y)^2 = [(\hat{Y} - E[\hat{Y}]) + (E[\hat{Y}] - Y)]^2$

Step 2: Expand the square

Use $(a+b)^2 = a^2 + 2ab + b^2$, we get:

$(\hat{Y} - Y)^2 = (\hat{Y} - E[\hat{Y}])^2 + 2(\hat{Y} - E[\hat{Y}])(E[\hat{Y}] - Y) + E[\hat{Y}] - Y)^2$

Step 3: Take the expectation.

Applying expectation $E[\cdot]$ to both sides:

$E[(\hat{Y} - Y)^2] = E[(\hat{Y} - E[\hat{Y}])^2] + 2E(\hat{Y} - E[\hat{Y}])(E[\hat{Y}] - Y)] + E[(E[\hat{Y}] - Y)$

Step 4: Simplify terms

1. The first term $E[(\hat{Y} - E[\hat{Y}])]^2$ is the variance of $\hat{Y}$, $Var(\hat{Y})$
2. The last term $E[(E[\hat{Y}] - Y)^2]$ is the square of the $(Bias(\hat{Y}))^2$
3. The middle term is: $2E[(\hat{Y} - E[\hat{Y}])(E[\hat{Y}] - Y)]$

Since $E[\hat{Y} - E[\hat{Y}]] = 0$, this term vanishes

Thus, we are left with:

$MSE = E[(\hat{Y} - Y)^2] = Var(\hat{Y}) + (Bias(\hat{Y}))^2$

MSE equals varians plus squared bias.