

Exercises

§5.4: Assessing a Regression Model

Python: There is an optional starter script that supports Python-based problem(s) on this sheet. Use it as scaffolding: read the comments, complete the lines marked **TODO**, and compare the output with the mathematics requested in the problem(s).

Problem 1 Suppose the residuals are $-2, 1, 0, 3$. Compute SSE, MSE, RMSE, and MAE.

Problem 2 Suppose $SSE = 24$ and $SST = 80$. Compute R^2 . Interpret the value relative to the mean-response baseline.

Problem 3 A model has $R^2 = -0.20$ on a test set. What does the negative value mean? Does it imply that the formula for R^2 is invalid?

Problem 4 Two models have the same MSE. Model A has many moderate residuals. Model B has nearly all residuals near zero but one very large residual. Why might the MSE alone be insufficient for choosing between them? What else would you inspect?

Problem 5 A residual plot shows a clear U-shaped pattern. What does this suggest about the linear model? Would simply running gradient descent for more iterations solve the problem? Explain.

Problem 6 A regression model has training RMSE = 1.2 and test RMSE = 7.5. What does the large difference suggest? Which value is more relevant to prediction on future unseen observations?

Problem 7 Use Python to calculate MSE, RMSE, MAE, and R^2 for

```
y = np.array([10., 12., 15., 18., 20.])  
y_pred = np.array([11., 11., 14., 20., 19.])
```

Then create a residual plot.