HGDS 200 - Foundations of Data-Driven Analysis 2018/2019 – Handout 4

1/14/2019 - 1/21/2019

Name: Connor Voglewede

1. The bottom-right data set would seem to exhibit a high degree of bias. It very poorly represent the asymptotic relationship in the data at x=8. It’s heavily skewed by a single outlier. It’s fixed by removing the outlier and re-calculating the linear regression.
2. None of the models seem to demonstrate a variance issue. There is no overfitting in any of the four models.
3. It’s not obvious to me which model has the lowest MSE. I wouldn’t be surprised if they were all equal.
4. Since n is equal for all of the models, I’d focus on comparing the sum of the squared residuals. To do that, I’d create literal squares for each point, with the length of a side being the value of the residual. All of the squares could be added together to show the sum of the squared residuals.