4.3: Nonparametric Regression

Dr. Bean - Stat 5100

Suppose each observation in my dataset had completely unique X-profiles. In other words, there was one and only one observation available for any combination of X-inputs. What might be some advantages and disadvantages for making the LOESS neighborhood so small that it fits each observation exactly?

- Advantage: The information contained in the sample is exactly preserved. This is often useful for spatial problems (called interpolation).

- Disadvantages:

    - Model fits have high variance (small changes to the observations included in the sample can cause large changes in the predicted values).
    - Models are over-fit (may not generalize well when predicting new observations)

Name a potential concern of using a LOESS model on a dataset with a very large number of explanatory variables.

As the number of dimensions increase, the distance between any two points becomes large. If the points are too spread out, the LOESS predictions can be highly sensitive to individual observations. High sensitivity can make these models fall prey to outlier values.

Name an example where the discrete predictions of a regression tree could be problematic.

(true story): Extreme value estimation requires continuous data. Dr. Bean used a regression tree to estimate data that was then used for extreme value estimation. The result was impractically large extreme estimates caused by the "jumps" in the input data.
(hypothetical): Predicting gas mileage of a car based on its features: it is problematic if a small change in a variable like car weight leads to a jump in the estimated gas mileage.