Class notes 5/10/2023

Discussion 1

What is the fundamental difference between parametric and nonparametric models?

- Parametric models (in context to regression) assume a specific functional form from between the response (Y) and the predictors (X).
- Y = f(X) + E(error)

Common misconception / confusing bit

- Previous course: A nonparametric model means there are fewer assumptions place on the model (typically the distribution of the data
  - Wilcoxon Rank Sum Test
  - Signed-Rank Test
  - Permutation Tests
- Note: In regression, we have nonparametric methods that deal with assumptions of normality as well as constant
- Just because a model is a complex(simple), it doesn't mean it is nonparametric(parametric).
- Key: Model governs how the predictions are made, not data

Discussion 2: Advertising Data Set

- Compare the MLR (multiple linearity regression) to fit additional KNN models changing k = 5 to k= 2, 10, 30. Summarize your findings.
  - As K increases, the fir looks more and more like the plane that MLR creates when adding no complexity
  - The MLR "looks like a piece of paper" it's a very simple and flat model.
  - K=30 is the closest to the original MLR meaning the higher the K the more simple it will be and vis versa when discussing the K is smaller

<u>Advice when learning from others code</u>

- Dissect: Break down code chunks line by line. If you do not know what the line does, print out the objects used as inputs, print out the output

Discussion 3

1. Using the provided graphs, do any of the predictor variables look like they are confounded with Sex? That is, is there any strong associated between Sex status and the other predictor variables.

The point of this question is to think about what confounding factors might look like from a scatterplot perspective and to learn a little about geom_smooth which is really handy.

Age Mildly confounded with Sex

- Much larger amount of Older women

Education Confounded with Sex

- Salaries look to trend up on average with education
- Graphics suggests that Males are more educated than Females
- Graphical anomaly. It looks like there is only 8 or 9 male observations. Are the blue points "behind the orange points?" Perhaps it is not a confounded as it first seems.

Definitely confounded, Males have higher relative percentage of education experience.

Discussion 3

- It is estimated that the baseline salaries of male employees is $765 higher than that of females who are the same age, have the same education, and experience.

- What does geom_smooth do?
  - It's a nonparametric regression technique.
  - Helpful to get a general idea of what f(x) looks like from your data. Can be helpful in EDA.
  - Cannot trust the ends (extrapolation).
  - Good for assessing interactions between categorical and numeric predictor.
  - Two predictors "interact" when the trend between the response and predictor changes depending on the remaining predictor.
  - This graph suggests no interaction is needed. The trend looks very similar indicating that the relationship between Salary and Age does not depend on Sex.

Discussion 4 Question 1

- Appears that an interaction is needed.
- Difference in OJ compared to VC appears to depend on the Dose level.
- Dos 0.5 and 1 (difference), and Dose 2 (no difference).

Discussion 4: Q2

- The overall F-test for the interaction effect (dose:supp) is statistically significant (p-value 0.0218).
- This supports the graphic investigated previously.