Hazel Ortiz

October 27, 2021

Week 9 Reading Questions

Note: I went to office hours with Anastasia and worked on some of these questions with her

1. Briefly (1 - 2 short paragraphs) describe at least two tradeoffs between the customized ML methods and the canned methods.

There are a few different tradeoffs when using the ML approach rather than the canned methods. From the reading, it seems that canned methods have some useful benefits against the ML approach. Canned methods have improved computational speed and stability which use faster algorithms that are less likely to encounter any numeric issues. Canned methods also have the benefit of stable definitions. Stable definitions simplify parameter estimation and involves smooth changes as parameters change. Models can also be fit onto scales that allow unconstrained optimization. Canned methods also have convention: when using a method like a linear regression, you don’t have to go into detail explaining what it is, and it may match better with models and parameters used in previous studies as well. With the ML approach, if it is used you need to explain everything carefully since it is a non-standard method. Building upon that last point, comparing hypothesis in canned methods can also be done easily in statistics programs such as R.

2. Briefly (1 - 2 sentences) describe each of the four key assumptions of the general linear modeling approach.

The four key assumptions are Normality of the residuals, Homogeneity, Fixed x, and independent observations. Normality of the residuals means data is normally distributed at x, homogeneity is constant variance along the range of x, Fixed x is when one is able to perfectly (without error) measure the value of variables, and Independent observations are observations that are independent of each other.

3. Explain how the normality assumption can be met in a general linear model, even if the response variable is not normally distributed. (1 - 2 paragraphs)

The normality assumption can be met in a general linear model by comparing the null hypothesis to the p-value and the model residuals. Looking at the example given, the data was tested with a normality test, the Shapiro-Wilk normality test showed us the p-value which was very small and smaller than the significance value, so the null was rejected, and it is not a normal distribution. However, looking at the next part of the example, more variables are added, and the residuals were tested, and the distribution was shown to be normal. So, even if the original response in not normally distributed, adding more variables from the data and testing residuals may show something different which may give us what is needed to meet the assumption for normality.