Thomas Courtney

Machine Learning, I

Professor Wilck

2/21/2021

I appreciated working with the wine dataset. This is a rather realistic task where it is important to predict the target one closest to zero. I enjoyed working with the different models and seeing how they provided different outputs with the same inputs.

First, I had to adjust my input data. Some of the data that we were originally given contained negative acidity, which is impossible. I changed Fixed Acidity, Volatile Acidity and Citric Acid so they would not contain negative numbers, I changed them to be the absolute value. For the rest of them, I changed NA's to be the mean of the rest of the data, as suggested by Professor Wilck.

I then created new values of the acidity levels and turned these into logs. Logs are great additions to machine learning models, as they are then placed to original scale.

I then ran the 6 suggested models. Out of all the models, the zero inflated Poisson model scored the highest. I then added in my additional variables and multiplied instead of added similar variables together. This returned a result of 3% of my test values to be less than .5, which I was happy to receive.