ST7016 Final Project Proposal

ZIHAO LI (u7031432)

For the final project, I'm going to use the dataset of Abalone on the UCI Machine Learning Website. The focus of the project is on the prediction of the age of abalone using the machine learning techniques. Normally, we can tell the age of abalone by observing their physical appearance (counting the rings on the shells). It's quite an old-fashioned and time-consuming way. What I'm going to do in the project is to explore the dataset of Abalone and do an explanatory data analysis to the dataset first. Then, investigating the relationships between our response variable 'rings' (indicates the age of abalone in this case) and each predictor (Other 8 attributes are sex, length, diameter, height, whole weight, shucked weight, viscera weight and shell weight). We are going to use all the predictors at the first stage and shrink the size of variables which are going to include in the final model during the later process. Currently, I have no clue about the appropriate model choice for our problem because the age of abalone is not fitted with any known distribution.

Questions/concerns about the project:

- I'm not sure my project argument is appropriate (e.g., Age prediction)

- By now, I have no clue about the prior distribution about my parameter of interest

- Is age prediction an interesting argument? Or maybe should I switch another question to focus on?

- Any advice for the direction that I should follow on the dataset of abalone