

Description: In these projects we will apply what we've learned in class to real-world datasets for analysis.

Tasks:

1. Using the **wine.csv** file provided in the datasets zip file, create a function to implement backward selection. Compare the results to best subset selection using the same metrics. Validate your models using a test/train split and compare on both the test and training data.
2. Create an ensemble model using bagging on the same dataset as above. Does it improve your metrics?
3. Using the **50_AdAgency.csv** file provided in the datasets zip file, create a function to implement a comparison of a one variable non-linear model profit predicted with one of the expense columns (your choice). You may use models of your choice, such as adjusting the number of knots in the spline model, adjusting the span in LOESS, changing the kernel or other parameters in the Gaussian process model, etc. Produce a graph of your metrics for at least three different metrics. Given these results, plot the best model. (This plot does not need to be built into the function, but should be available from the results.)

Submission: Write up a work document describing your code and the results obtained. Include with your submission your R code file.