

Instructions: For these weekly coding assignments, you will be asked to extend the examples from class to create custom code to answer the questions below. You will create an R code file that uses built-in datasets as the data sources. You will write the code, and an example showing that the code works. Be sure to include any packages in the code that are required for the functions to run (you may want to clear the environment in RStudio before your final check to make sure nothing is missing). The instructor will run the file to ensure that it works with no errors. Clearly label your code so it's clear which question/task is being responded to.

Submission:

A word document with any explanations (if needed), and a clearly labeled R code file.

Tasks/Questions:

1. Modify the bagging example from lecture to also use random subsets selections of predictor variables of a given size, say $k=3$.

2. Apply the boosting algorithm to BostonHousing dataset from the package mlbench.

```
install.packages("mlbench")  
library(mlbench)  
data(BostonHousing)
```

This dataset has more than 500 observations. Create a test and training set for validation. Test appropriate metrics and visualizations.

3. Create a stacked algorithm, following and expanding the example from lecture, that includes polynomials up to degree 6, includes 5 span settings for loess, splines, and Gaussian processes using different kernels or scaling dimensions (you can use different GP packages or code it up from scratch as we did in previous lectures, or use the same package if it allows different kernel and/or scale selections). Output metrics and graphs to test the performance. Use the abalone dataset from the mlbench package. I suggest creating a correlation analysis first to select the best single variable to use in your predictions (you may want to create graphs to see if any relationships might be nonlinear).