STA 561: Homework 3 (Due Feb 24 at midnight)

Reminder: work together! Share ideas, brainstorm, explain/verify your answers but write up your own work. Your homework should be submitted as pdf file generated using either latex or an python notebook.

1. (A programming problem) In this homework, you will write a python or R function that automatically tunes blackbox regression models. Your function should take as input:

   - A learning algorithm, i.e., a function that takes as input a matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ and a vector of responses $\mathbf{Y} \in \mathbb{R}^n$ and returns a function that maps inputs to outputs, i.e, maps $\mathbb{R}^p$ into $\mathbb{R}$
   - Training data $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ and $\mathbf{Y} \in \mathbb{R}^n$
   - A regularization method that belongs to the set $\{\text{Dropout}, \text{NoiseAddition}, \text{Robust}\}$
   - An positive integer M indicating the number of Monte Carlo replicates to be used if the method specified is Dropout or NoiseAddition
   - A vector $\mathbf{c}$ of column bounds to be used if the method specified is Robust
   - A positive integer K indicating the number of CV-folds to be used to tune the amount of regularization, e.g., $K = 5$ indicates five-fold CV
   - A criterion to be used to evaluate the method that belongs to the set $\{\text{MSE}, \text{MAD}\}$ where MSE encodes mean square error and MAD encodes mean absolute deviation.

   Your function should output a predictive model that optimizes the specified criterion using the specified method. I.e., if the function were called with Dropout, $M = 100$, $K = 10$ and MSE it would find the amount of dropout (with $M = 100$ random dropout vectors per observation) that minimizes the ten-fold cross-validated MSE and return the learning algorithm trained with this amount of dropout.

   This homework is worth 20 points (twice that of normal homework) but I'll score it out of 10 so there's opportunity for extra points here. Grading will be based on: correctness, readability, style, and documentation.