STA 561: Homework 3 (Due Feb 21 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 story problem) *Background: I recently attended a teaching workshop as part of my new faculty onboarding. At this workshop, I learned that story problems are a great way for students to learn by connecting class materials to their everyday lives. Here we go, teacher training in action!*

   Tommy took a long drag on his cigarette, taking it down to its filter. He let the butt fall lazily from his fingers and ground it under his thick-soled hiking boot. It was an old habit, there certainly wasn't any risk of starting a fire in this godforsaken sandpile. He had spent the last four months combing through the White Dessert following a smeared and ratty treasure map that he'd found in a foreclosed storage locker that he'd won at auction. The journey hadn't been an easy one.

   To begin with, he'd started his quest without much of a plan. After finding the map, high on the prospect of what he called 'Go Tell Aunt Rhodie' riches, he'd occupied the weeks leading up to his departure telling people what he'd really thought of them (mostly that they were unambitious garbage people that would never amount to anything and would regret not being nicer to him when he was richer than Moses). Most of the people who knew him well were unfazed by Tommy's insults and few bothered to correct his "richer than Moses" comment by reminding him that Moses had gone from being a prince to being a shepherd. After burning his bridges, Tommy stopped by a local outdoor outfitters where be bought a canteen, a knock-off Swiss Army knife, hiking boots, a tent, sunscreen, a compass, a few changes of clothes, and a case of emergency food rations. He took an unlicensed cab to the airport (without having purchased an airline ticket). Three days later and one box of emergency food rations later, he touched down in Cairo ready to find his fortune.

   He hired a local man that he found lingering outside the airport to be his guide. By nightfall they were camping in the White Dessert, sleeping under the stars and alongside their rented camels. Tommy's tour guide, Tarek, it turned out, knew about as much as Tommy when it came to survival in the dessert—which is to say, nothing. After three months of wandering they had run out of food and discipline and soon

were trotting aimlessly through the dessert until their mounts refused to go on. That's when they tried to eat the camels. Tommy had attempted to take them with the faux Swiss army knife but quickly found it was both too small and too dull to do the job properly. He considered trying to bludgeon the camels with rocks but he didn't have the strength. That's when his eyes fell on Tarek.

The fight was an ugly and heartbreaking affair between two weak, desperate, and unskilled fighters. But ultimately, small and dull though it was, the knife was the difference maker. It had taken much more time, effort, and agony to finish the job than Tommy had expected. When it was finally over, when the gurgles and spasms had ceased, Tommy studied Tarek's body. It was coated in blood and sweat and sand. Tommy felt bile rising his throat but knew he mustn't allow himself to become sick, the dehydration would be fatal. He turned and walked away from the gruesome scene. He'd just need to gather his wit, then he'd return and eat for the first time in weeks. Or, so he'd thought.

After only a few minutes of walking, he had found *it*. The treasure map had an etching of a cave at the foot of small stone obelisk—an obelisk that he was now starring at. Fearing it might be a hallucination, he rubbed his eyes, which had become gritty and dry to match his surroundings. But the obelisk remained. It was real. He pulled a cigarette from his breast pocket (cigarettes had been the only item he had successfully rationed). Starring at the obelisk he thought about the potential riches inside.

Tommy found the cave entrance easily, it was exactly as depicted in the etching. He had expected poison darts, malicious mummies, or some other kind of resistance, but he found his way through the cave easily. The cave formed a smooth tunnel of uniform diameter leading straight into the earth. The tunnel was so uniform that Tommy wondered if it had been man-made like the obelisk. Lighting his way with his cigarette lighter, he followed the tunnel for more than half an hour before he finally spotted a light ahead. He quickened his step, seeing in his mind's eye, the new life these riches would afford him.

The light source came into view, it was a computer screen. Sitting on an ornate stone pedestal was a Lenovo Yoga 9 touch screen laptop computer. The desktop showed a single file: `PredictoTron2022.exe`. "Jackpot." Tommy said, his voice filled with triumph and anticipa-

tion. He double-clicked the file and simple graphical user interface appeared on the screen. There was a place to enter a file containing predictors and another for file containing responses. Given these inputs, the PredictoTron2022 algorithm would create a predictive model with unsurpassed performance. But, as is the case with any treasure found at the end of a perilous journey plagued by murder of friend and camel, there was a catch. The model needed to be tuned. Without proper tuning the model would overfit the data leading to low-quality predictions. Because Tommy had never taken STA561, he had no idea how to regularize a blackbox model. The PredictoTron2022 was useless to him. He left cave a despondent and broken shell of a person. Shortly after, he was eaten by his camels.

Your job is to 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.