Homework 3 - Decision Trees & Recursion

# Instructions

Most homework assignments for this class will consist of two tasks:

1. **Reading/Video**: This will be an article or video about a topic in machine learning related to what we have learned so far in the class. After you read the article or watch the video, we will ask you to answer some comprehension questions (typed or handwritten responses are fine), which will relate to the class content, coding activity, and/or to the reading/video. We will start off class next week by discussing these comprehension questions (remind us if we forget!)
2. **Coding**: This consists of completing the coding file you were working on in class today. This does not include the bonus challenge, although you are welcome to complete that if you want!

# Reading

Please read the below article and the two videos below about a new electrical grid system. The first two questions below are just a continuation of last week’s theme of Decision Trees and this week’s coding assignment. The following two questions are about the reading.

Tesla is connecting 50000 homes in Australia to create a power grid:

https://thenextweb.com/au/2018/02/05/tesla-is-connecting-50000-homes-in-australia-to-create-a-power-grid/

## Comprehension Questions

1. Pick two of the below questions (part a through e) to answer you did not answer last week.

* 1. As you saw, the accuracy of our decision tree is around 50%. That is as bad as if we tossed a coin and predicted “UP/DOWN”! What does this tell you about making stock market predictions using historical data? Do you have any ideas for other data we could take into account to make a better predictor?
  2. All of our training data came from 2016 and all of our test data came from 2017. What are the pros/cons of doing this?
  3. When splitting along the first attribute (simple moving average), we try every possible split. However, the number of splits is very large (equivalent to the size of the training data minus one). Do you have any ideas for a faster way to evaluate splits, that may not get us the best split but will get us a good enough split (there are many options).
  4. What if instead of each split being binary (i.e. yes or no), some had more options. For example, if instead of UP/DOWN for the second attribute we had UP/DOWN/SAME, then we may have to have 3-way splits. What parts of the code would we have to modify to account for this?
  5. Decision trees in general are useful in day-to-day-life. Fundamentally, they are just a way of organizing cascading conditions (i.e. if I do X, then I will go down this route of the tree). Think of ways in which decision trees might be useful in your life, or the lives of people you know.

1. In the reading about Tesla, it mentions that Tesla is decentralizing a power grid. In fact machine learning was used in this process. How do you think machine learning would help in re-designing and decentralizing a power-grid?
2. On a broad level, how would you start thinking about optimizing a power grid?