CS405 Machine Learning
Lab #1 Decision Trees

Pre-Lab (20 points):

1. Familiarize yourself with the following materials including those equations involved:
   1) Decision tree: https://en.wikipedia.org/wiki/Decision_tree
   2) ID3 algorithm: https://en.wikipedia.org/wiki/ID3_algorithm
   3) C4.5 algorithm: https://en.wikipedia.org/wiki/C4.5_algorithm

   Think over and answer the follow questions:
   1) what are the advantages and disadvantages of decision tree algorithms?
   2) what kind of problems could be solved by decision trees?
   3) how to generate a decision tree and how to overcome the over-fit problem?
   4) what is the computational complexity of the decision tree algorithm.

2. Read materials about random forest, such as:
   https://en.wikipedia.org/wiki/Random_forest
   please summarize the difference between random forest and decision tree.

3. When the features are continuous instead of discrete, how to select cut-points in decision trees?

4. Could a tree model be used for regression? How to implement the corresponding algorithm? Please compare two cases of regression and classification by using tree models.