

Assignment 1 : Decision trees

Your goal in this assignment is to implement a Decision Tree Classifier on the adult dataset. Link: <http://archive.ics.uci.edu/ml/machine-learning-databases/adult/>

You can use any programming language to implement your decision tree. (python is preferable)

You have to do the following tasks in this assignments.

- 1) Download the dataset and perform statistical analysis on the given attributes which includes calculations like mean, mode, median, min, max and standard deviation etc. describe what are your intuitions for this dataset.
- 2) Split the dataset into training, validation and testing sets. (preferably Training = 60 , validation 30, testing =10). Discuss which technique was used to split the dataset and why?
- 3) Train your decision tree with different variations of hyper parameters like depth, no of trees, splitting criteria (e.g. entropy or information gain) etc. and perform thorough empirical analysis of which performance parameters give the best results. Also perform cross validation to validate your design choices.
- 4) Write Decision rules for every tree and also the confusion matrix. And after successful parameter selection test your model only once on testing data.
- 5) Report the best results for your trained Decision tree classifier in a report. Report should be short and to the point. Don't copy paste screen shots or dumb information on pages, rather make tables and make it presentable.

(* You can not use any custom libraries for this assignment you must code each and every part of the assignment yourself otherwise a straight zero will be awarded. Plagiarism in any part of the assignment is academic dishonesty and is a BIG NO NO !

* You must apply all the learnt concepts in the class not just the ones mentioned in the assignment)

BEST OF LUCK !!