

ComS 573 Machine Learning

Lab 3 (100 points)

Decision Tree Learning

Note

- For help with laboratory assignments, please contact TA.
- You may perform the experiments using a machine learning package that has implemented (some version of) decision tree learning algorithm, such as Weka (<https://www.cs.waikato.ac.nz/ml/weka/>), etc.

In this lab assignment, you will experiment with the Decision Tree classifier.

1 Dataset

We will use the following data set from the UC Irvine Machine Learning Repository:

- Congressional Voting Records Data Set at (<https://archive.ics.uci.edu/ml/datasets/Congressional+Voting+Records>). Note the dataset has missing values.

2 Tasks

1. Learn a Decision Tree classifiers on the data set (for example, using J48 in Weka). Visualize the tree constructed by the decision tree algorithm. (Food for thought: Are there some interesting rules that make sense based on what you understand about the data?)
2. Report the accuracy of the Decision Tree classifier using 5-fold cross-validation (Most of ML packages including Weka have utility function for performing Cross-validation). Report 95% confidence interval.
3. Perform the following experiments to study the stability of decision tree learning algorithm over the variability of data samples.
 - (a) Randomly split the dataset into 5 data sets of (roughly) equal size D_1, D_2, \dots, D_5 .
 - (b) For $i = 1, 2, \dots, 5$, each time use D_i as test data and the rest as training data to learn a decision tree and measure its accuracy p_i .
 - (c) Visualize the five trees constructed. Do the five trees differ with each other and with the tree constructed using all the data (in Task 1)? How much do their accuracies differ?

3 What to turn in

Turn in via Canvas the following:

- A lab report (in pdf file) with your experimental results.
- A Readme file with instructions on how to reproduce your experiments. You should specify the parameters of every experiment in a way such that they can be replicated by the TA.
- Any source code that you may have written (in a zip file).