

# Assignment 1: Decision Tree Induction

*09/22/2014*

- Class: CS 4375 - Introduction to Machine Learning
- Instructor: Dr. Vincent Ng
- Semester: Fall 2014
- School: The University of Texas at Dallas

## Contents

There are two folders in this repository, each for one part of the assignment:

- `part_1`: Written Problems
- `part_2`: Programming Problem
  - `part_2/cond_entropy.R`: Compute conditional entropy value
  - `part_2/Tree.R`: Define ID3 Tree class
  - `part_2/learn_tree.R`: Learn a ID3 Tree
  - `part_2/ID3_decision_tree.R`: Run using command line with two parameters, `<TRAIN_FILE>` and `<TEST_FILE>`

## Run

```
Rscript ID3_decision_tree.R <TRAIN_FILE> <TEST_FILE>
```

## Platforms

- Windows 8.1
- Ubuntu 14.04

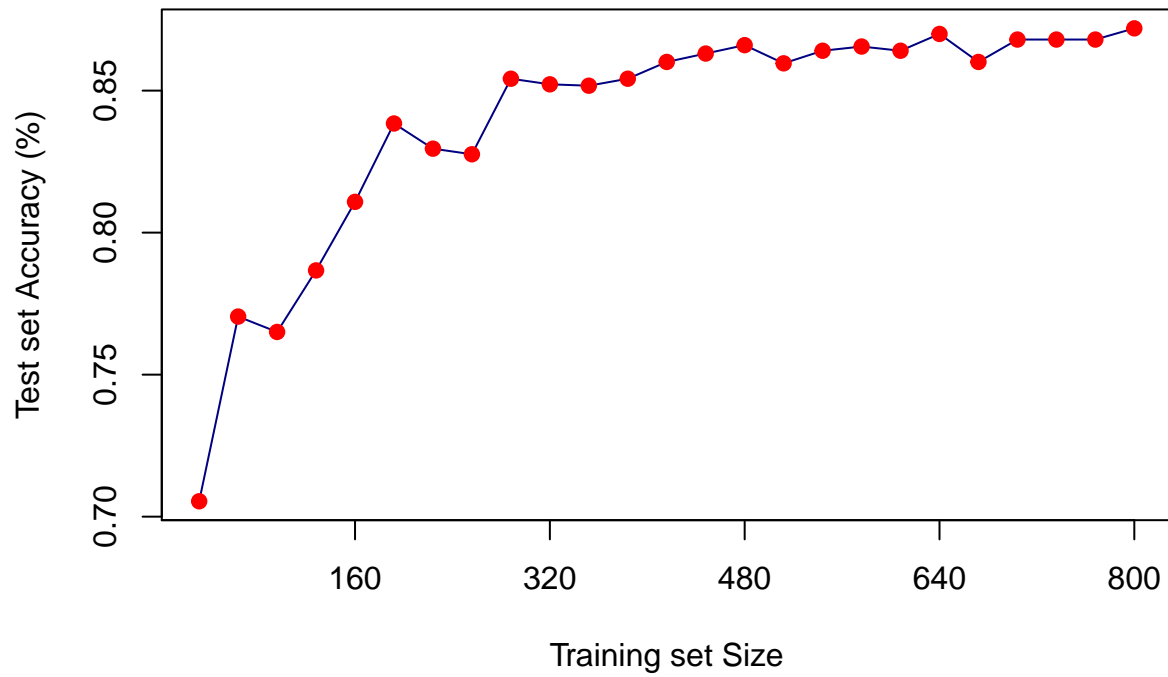
## Contributors

- Isaac Butterfield
- Tuan Anh Pham

## Learning Curve

```
source("learning_curve.R")
```

## Learning Curve of ID3 Decision Tree



Test set accuracy increases as size of training set increases. However, after training set size exceeds a value (about 480), test set accuracy does not have significant change. This pattern represents the usual properties of a learning curve.