

## Homework1 of Machine Learning

### Part1, readme

```
# README for Homework1 of CS6375.502
# In root directory of Homework1, use "./configure;make -k" to compile

./configure;make -k
```

# Then, run the program by using following instructions

```
src/main 100 15 data_sets1/training_set.csv data_sets1/validation_set.csv data_sets1/test_set.csv no
src/main 100 15 data_sets1/training_set.csv data_sets1/validation_set.csv data_sets1/test_set.csv yes
src/main 100 15 data_sets2/training_set.csv data_sets2/validation_set.csv data_sets2/test_set.csv no
src/main 100 15 data_sets2/training_set.csv data_sets2/validation_set.csv data_sets2/test_set.csv yes
```

### Part2, accuracy on datasets

	Algorithm 1	Algorithm 2
Test set 1	75.15%	75.20%
Test set 2	74.50%	74.17%

Accuracies on the test set for decision trees constructed using the two heuristics.

By choosing L and K both from 1 to 100, run test on testset1 and testset2 both 10,000 times, we choose 5 combinations which makes Algorithm1 on testset1 five salient values for different accuracy parts, and similarly choose another 5 combinations for Algorithm2's accuracy on testset1. After that, we just use the 10 combinations of L and K and get the accuracy on testset2. Below is the accuracy table.

	Algorithm1	Algorithm2		Algorithm1	Algorithm2	L	K
Test set1	73.35%	75.20%	Test set2	75.83%	76.00%	68	4
Test set1	75.00%	75.50%	Test set2	72.67%	75.17%	41	58
Test set1	75.35%	75.85%	Test set2	74.50%	74.50%	10	30
Test set1	75.70%	74.90%	Test set2	74.50%	74.17%	44	76
Test set1	76.75%	74.95%	Test set2	74.83%	75.33%	49	87
Test set1	75.25%	73.40%	Test set2	75.17%	75.33%	35	11
Test set1	75.55%	75.00%	Test set2	75.50%	76.33%	22	45
Test set1	75.85%	75.35%	Test set2	74.33%	74.83%	47	18
Test set1	75.40%	75.70%	Test set2	74.50%	75.00%	4	36
Test set1	75.85%	76.80%	Test set2	74.50%	74.83%	91	94

Accuracies on the test set for post-pruned decision trees constructed using the two heuristics.