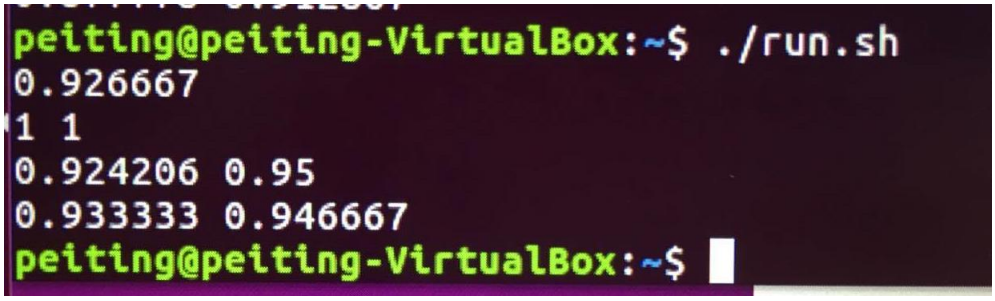
Hw1 Report

Environment: Ubuntu 16.04.3 LTS
Language: c++
Using library: <cstdio> <iostream> <vector> <cmath> <algorithm> <time.h> <stdlib.h> <cstring>
Result:



Explain:
Use struct dat to be the datatype of 1 instance
iris[0] save sepal len iris[1] save sepal wid
iris[2] save petal len iris[3] save petal wid
iris[4] save class as float
use vector dataset to save all the instances
class ID3 : use to present remaining dataset
calculate bound set of [n] attribute in dataset
calculate information gain of [n] attribute with bound
class dctree : build decision tree (tree,set)
set : choose the best attribute and bound by information gain, save tree->node = attribute tree->value = bound
split dataset <bound go to leftsubset
>=bound go to rightsubset
Build left tree by left subset (recursive)
Build right tree by right subset (recursive)

testdata : test instance on tree 用 tree->node,value 決定判

斷的 attribute>=bound go right else left

shuffle dataset to randdata
run K-fold :
every time (total 5 times/ i:0-4)

{
vector test push group i of instances(30 instances) while vector train push the other 120 instances from randdata
use train data to buildtree
test each instance of testdata to the tree which built by train data
calculate accuracy[i](predict right/total)
class0 precision[i](predict right in class0/actual num of class0)
class0 recall[i](predict right in class0/predict num of class0)
so do class1,2
}

Get average accuracy recall precision by 相加/5

=>print
[Total accuracy]
[Precision of class 0] [Recall of class 0]
[Precision of class 1] [Recall of class 1]
[Precision of class 2] [Recall of class 2]

run.sh:
#!/bin/bash
G++ 0416043_hw1.cpp -o fout //compile the source file

./fout //執行