

Name: Naveenraj Palanisamy

NetId: NXP154130

Machine Learning Assignment -1 (Decision Tree)








(Note: Runs in python 3.5)

(Note: File Path should be full path)




Zip file name: Naveenraj_NXP154130_ML_Assignment_1.zip

Folder Name: Naveenraj_NXP154130_ML_Assignment_1




Folder Structure:

	data_sets1	2/13/2016 12:04 A...	File folder	
	data_sets2	2/13/2016 12:04 A...	File folder	
	ID3-purning tree for each itteration	2/13/2016 12:04 A...	File folder	
	Trees before and after pruning	2/13/2016 12:04 A...	File folder	
	Variance-impurity purning each itterat...	2/13/2016 12:04 A...	File folder	
	ML_Refinement	2/12/2016 10:12 PM	JetBrains PyCharm	26 KB
	Report_NXP154130	2/12/2016 11:11 PM	Microsoft Word D...	21 KB
























Data_set1-> Contains first set of given data

	test_set	2/13/2016 12:04 A...
	training_set	2/13/2016 12:04 A...
	validation_set	2/13/2016 12:04 A...

Data_set2->Contains second set of given data




















	test_set	2/13/2016 12:04 A...
	training_set	2/13/2016 12:04 A...
	validation_set	2/13/2016 12:04 A...

ID3-purning tree for each iteration-> Contains all tree got during pruning for ID3 for each L (L is given as 50)





	ID30I	2/12/2016 9:52 PM	DAT File	8 KB
<input checked="" type="checkbox"/>	ID31I	2/12/2016 9:52 PM	DAT File	8 KB
	ID32I	2/12/2016 9:52 PM	DAT File	8 KB
<input type="checkbox"/>	ID33I	2/12/2016 9:52 PM	DAT File	8 KB
	ID34I	2/12/2016 9:52 PM	DAT File	8 KB
	ID35I	2/12/2016 9:52 PM	DAT File	8 KB
	ID36I	2/12/2016 9:52 PM	DAT File	8 KB
	ID37I	2/12/2016 9:52 PM	DAT File	8 KB
	ID38I	2/12/2016 9:52 PM	DAT File	8 KB
	ID39I	2/12/2016 9:52 PM	DAT File	8 KB
	ID310I	2/12/2016 9:52 PM	DAT File	8 KB
	ID311I	2/12/2016 9:52 PM	DAT File	8 KB
	ID312I	2/12/2016 9:52 PM	DAT File	8 KB
	ID313I	2/12/2016 9:52 PM	DAT File	8 KB
	ID314I	2/12/2016 9:52 PM	DAT File	8 KB
	ID315I	2/12/2016 9:52 PM	DAT File	8 KB
	ID316I	2/12/2016 9:52 PM	DAT File	8 KB
	ID317I	2/12/2016 9:52 PM	DAT File	8 KB
	ID318I	2/12/2016 9:52 PM	DAT File	8 KB
	ID319I	2/12/2016 9:52 PM	DAT File	8 KB
	ID320I	2/12/2016 9:52 PM	DAT File	8 KB
	ID321I	2/12/2016 9:52 PM	DAT File	8 KB
	ID322I	2/12/2016 9:52 PM	DAT File	8 KB
	ID323I	2/12/2016 9:52 PM	DAT File	8 KB
	ID324I	2/12/2016 9:52 PM	DAT File	8 KB

Type: DAT File
Size: 7.89 KB
Date modified: 2/12/2016 9:52 PM

Variance-impurity purning each iteration -> Contains all tree got during pruning for Impurity variance for each L (L is given as 50)

	VARIENCE0I	2/12/2016 9:52 PM	DAT File	10 KB
	VARIENCE1I	2/12/2016 9:52 PM	DAT File	10 KB
	VARIENCE2I	2/12/2016 9:52 PM	DAT File	10 KB
	VARIENCE3I	2/12/2016 9:52 PM	DAT File	10 KB
	VARIENCE4I	2/12/2016 9:52 PM	DAT File	10 KB
<input type="checkbox"/>	VARIENCE5I	2/12/2016 9:52 PM	DAT File	10 KB
	VARIENCE6I	2/12/2016 9:52 PM	DAT File	10 KB
	VARIENCE7I	2/12/2016 9:52 PM	DAT File	10 KB
	VARIENCE8I	2/12/2016 9:52 PM	DAT File	10 KB
	VARIENCE9I	2/12/2016 9:52 PM	DAT File	10 KB
	VARIENCE10I	2/12/2016 9:52 PM	DAT File	10 KB
	VARIENCE11I	2/12/2016 9:52 PM	DAT File	10 KB
	VARIENCE12I	2/12/2016 9:52 PM	DAT File	10 KB
	VARIENCE13I	2/12/2016 9:52 PM	DAT File	10 KB
	VARIENCE14I	2/12/2016 9:52 PM	DAT File	10 KB
	VARIENCE15I	2/12/2016 9:52 PM	DAT File	10 KB
	VARIENCE16I	2/12/2016 9:52 PM	DAT File	10 KB
	VARIENCE17I	2/12/2016 9:52 PM	DAT File	10 KB
	VARIENCE18I	2/12/2016 9:52 PM	DAT File	10 KB
	VARIENCE19I	2/12/2016 9:52 PM	DAT File	10 KB

Trees before and after pruning-> Contains tree for ID3 and Impurity Variance (Both before and after pruning)

	ID3 after pruning	2/12/2016 9:52 PM	DAT File	7 KB
	ID3 before pruning	2/12/2016 9:52 PM	DAT File	8 KB
	VARIENCE AFTER PRUNING	2/12/2016 9:52 PM	DAT File	10 KB
	VARIENCE BEFORE PRUNING	2/12/2016 9:52 PM	DAT File	10 KB

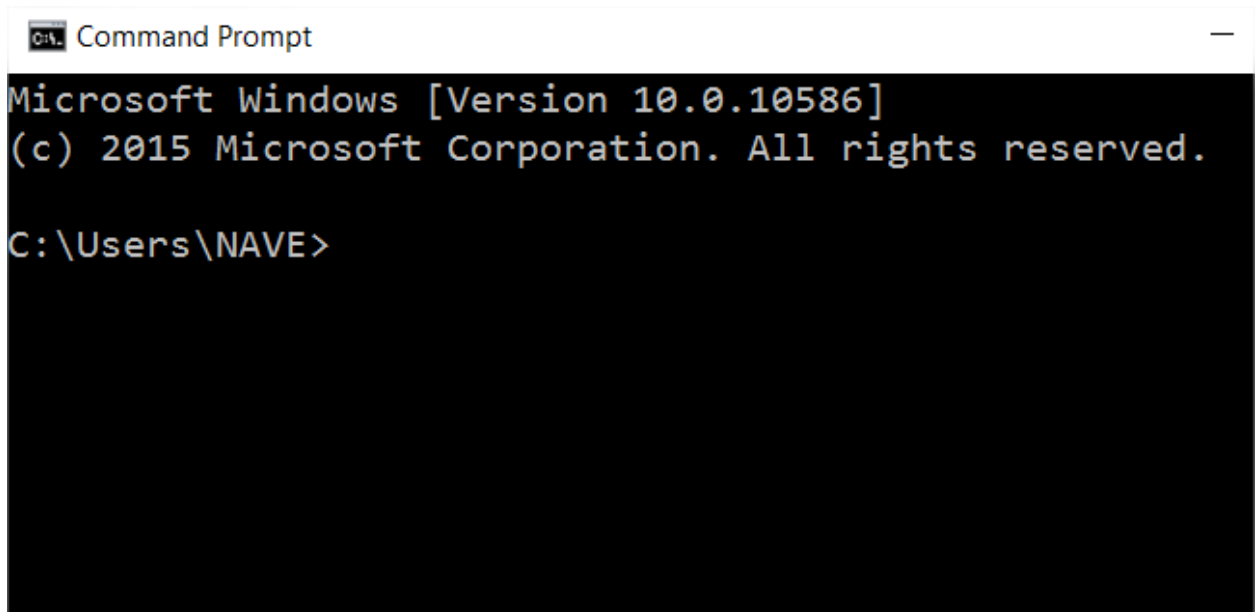
ML_Refinement.py -> python program that need to run.

Running the program steps: ->

- 1) Go to 'command prompt'
- 2) Go to folder where python is installed.
- 3) Given command as (python.exe 'full path to-> ML_Refinement.py' filelocation1 filelocation2 filelocation3 K L YES/NO)
- 4) Giving YES to last command line argument will display the tree in command prompt and also prints Trees before and after in the current path.
- 5) Giving NO will only prints Trees before and after in the current path.

Sample Run:

- 1) Start -> cmd



```
Command Prompt
Microsoft Windows [Version 10.0.10586]
(c) 2015 Microsoft Corporation. All rights reserved.

C:\Users\NAVE>
```

- 2) Copy the path where python 3.5 is installed.
For me path is:
C:\Users\NAVE\AppData\Local\Programs\Python\Python35-32\python.exe
- 3) Copy the path where folder is downloaded.
For me it is:
C:\Naveenraj_NXP154130_ML_Assignment_1\ML_Refinement.py
- 4) Copy the file path for all the files in the data set
For me it is: (in the order of training, validation and test)
First ->
C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\training_set.csv
Second->
C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\validation_set.csv
Third ->
C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\test_set.csv
- 5) Decide values for L and K (I gave as 20 4)

6) Decide need to print tree or not (I gave YES)

Now run the program with all the above mentioned values.

C:\Users\NAVE\AppData\Local\Programs\Python\Python35-32\python.exe

C:\Naveenraj_NXP154130_ML_Assignment_1\ML_Refinement.py

C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\training_set.csv

C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\validation_set.csv

C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\test_set.csv 20 4 YES

```
C:\Users\NAVE>C:\Users\NAVE\AppData\Local\Programs\Python\Python3
5-32\python.exe C:\Naveenraj_NXP154130_ML_Assignment_1\ML_Refinem
ent.py C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\training
_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\valida
tion_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\te
st_set.csv 30 40 YES_
```

```
XD : 1
XT : 1
XN : 1
XE : 0
XO : 1
XO : 0
XE : 1
XN : 0
XM : 1
XO : 0
XC : 1
XF : 0
XF : 1
XB : 0
XB : 1
XC : 0
XO : 1
ID3 before pruning 74.5
ID3 after pruning 75.5
variance impurty before pruning 66.17
variance impurty after pruning 66.33
```

nt_1\ML_

l test)

data_sets2

data_sets2

data_sets2

es.

Different try and results are given below:

```
C:\Users\NAVE>C:\Users\NAVE\AppData\Local\Programs\Python\Python35-32\python.exe C:\Naveenraj_NXP154130_ML_Assignment_1\ML_Refinement.py C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\training_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\validation_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\test_set.csv 20 4 NO
ID3 before pruning 74.5
ID3 after pruning 74.83
variance impurty before pruning 66.17
variance impurty after pruning 66.33
```

```
C:\Users\NAVE>C:\Users\NAVE\AppData\Local\Programs\Python\Python35-32\python.exe C:\Naveenraj_NXP154130_ML_Assignment_1\ML_Refinement.py C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\training_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\validation_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\test_set.csv 10 14 NO
ID3 before pruning 74.5
ID3 after pruning 74.5
variance impurty before pruning 66.17
variance impurty after pruning 66.17
```

```
C:\Users\NAVE>C:\Users\NAVE\AppData\Local\Programs\Python\Python35-32\python.exe C:\Naveenraj_NXP154130_ML_Assignment_1\ML_Refinement.py C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\training_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\validation_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\test_set.csv 19 9 NO
ID3 before pruning 74.5
ID3 after pruning 75.33
variance impurty before pruning 66.17
variance impurty after pruning 66.5
```

```
C:\Users\NAVE>C:\Users\NAVE\AppData\Local\Programs\Python\Python3
5-32\python.exe C:\Naveenraj_NXP154130_ML_Assignment_1\ML_Refinem
ent.py C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\training
_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\valida
tion_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\te
st_set.csv 190 9 NO
ID3 before pruning 74.5
ID3 after pruning 75.83
variance impurty before pruning 66.17
variance impurty after pruning 67.33
```

```
C:\Users\NAVE>C:\Users\NAVE\AppData\Local\Programs\Python\Python3
5-32\python.exe C:\Naveenraj_NXP154130_ML_Assignment_1\ML_Refinem
ent.py C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\training
_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\valida
tion_set.csv C:\Naveenraj_NXP154130_ML_Assignment_1\data_sets2\te
st_set.csv 190 190 NO
ID3 before pruning 74.5
ID3 after pruning 74.83
variance impurty before pruning 66.17
variance impurty after pruning 68.83
```

ID3 tree after pruning will look like (for one random condition):

XI: 1

| XK: 0

| | XC: 0

| | | XS: 0

| | | | XG: 1

| | | | | XO: 1

| | | | | | XL: 0

| | | | | | | XH: 0 1 (goes on)(..... to file the full file see in the trees before and after folder)