Decision Tree

February 19, 2020

1 Introduction

Decision tree is one of the many machine learning algorithms. It is used to classfier: given input data, it is class A or Class B? We can visualize a decision tree using the Python module pydotplus and the module graphviz.

```
In [2]: import pandas as pd
```

```
In [9]: myData=pd.read_csv("F:\\Learning Resources Jan 2020\\JKUAT MAIN\\AI\\vertebrate\\vertel
```

```
In [10]: myData
```

Out[10]:

	Name	Warm-blooded	Gives Birth	Aquatic Creature	Aerial Creature	Has Legs	Hibernates	Class
0	human	1	1	0	0	1	0	mammals
1	python	0	0	0	0	0	1	reptiles
2	salmon	0	0	1	0	0	0	fishes
3	whale	1	1	1	0	0	0	mammals
4	frog	0	0	1	0	1	1	amphibians
5	komodo	0	0	0	0	1	0	reptiles
6	bat	1	1	0	1	1	1	mammals
7	pigeon	1	0	0	1	1	0	birds
8	cat	1	1	0	0	1	0	mammals
9	leopard shark	0	1	1	0	0	0	fishes
10	turtle	0	0	1	0	1	0	reptiles
11	penguin	1	0	1	0	1	0	birds
12	porcupine	1	1	0	0	1	1	mammals
13	eel	0	0	1	0	0	0	fishes
14	salamander	0	0	1	0	1	1	amphibians

```
8                 0         1           0     mammals
9                 0         0           0      fishes
10                0         1           0    reptiles
11                0         1           0       birds
12                0         1           1     mammals
13                0         0           0      fishes
14                0         1           1  amphibians
```

```
In [11]: myData['Class']=myData['Class'].replace(['fishes','birds','amphibians','reptiles'],'n
```

```
In [12]: myData
```

Out[12]:

	Name	Warm-blooded	Gives Birth	Aquatic Creature	\
0	human	1	1	0	
1	python	0	0	0	
2	salmon	0	0	1	
3	whale	1	1	1	
4	frog	0	0	1	
5	komodo	0	0	0	
6	bat	1	1	0	
7	pigeon	1	0	0	
8	cat	1	1	0	
9	leopard shark	0	1	1	
10	turtle	0	0	1	
11	penguin	1	0	1	
12	porcupine	1	1	0	
13	eel	0	0	1	
14	salamander	0	0	1	

	Aerial Creature	Has Legs	Hibernates	Class
0	0	1	0	mammals
1	0	0	1	non-mammals
2	0	0	0	non-mammals
3	0	0	0	mammals
4	0	1	1	non-mammals
5	0	1	0	non-mammals
6	1	1	1	mammals
7	1	1	0	non-mammals
8	0	1	0	mammals
9	0	0	0	non-mammals
10	0	1	0	non-mammals
11	0	1	0	non-mammals
12	0	1	1	mammals
13	0	0	0	non-mammals
14	0	1	1	non-mammals

```
In [19]: pd.crosstab([myData['Warm-blooded'],myData['Gives Birth']],myData['Class'])
```

Out[19]:

Class		mammals	non-mammals
Warm-blooded	Gives Birth		

```
         0            0                 0            7
                      1                 0            1
         1            0                 0            2
                      1                 5            0
```

```
In [20]: from sklearn import tree
         y=myData['Class']
         x=myData.drop(['Name','Class'],axis=1)
```

```
In [18]: y
```

```
Out[18]: 0         mammals
         1     non-mammals
         2     non-mammals
         3         mammals
         4     non-mammals
         5     non-mammals
         6         mammals
         7     non-mammals
         8         mammals
         9     non-mammals
         10    non-mammals
         11    non-mammals
         12        mammals
         13    non-mammals
         14    non-mammals
         Name: Class, dtype: object
```

```
In [21]: x
```

```
Out[21]:     Warm-blooded  Gives Birth  Aquatic Creature  Aerial Creature  Has Legs  \
         0               1            1                 0                0         1
         1               0            0                 0                0         0
         2               0            0                 1                0         0
         3               1            1                 1                0         0
         4               0            0                 1                0         1
         5               0            0                 0                0         1
         6               1            1                 0                1         1
         7               1            0                 0                1         1
         8               1            1                 0                0         1
         9               0            1                 1                0         0
         10              0            0                 1                0         1
         11              1            0                 1                0         1
         12              1            1                 0                0         1
         13              0            0                 1                0         0
         14              0            0                 1                0         1

             Hibernates
         0            0
```

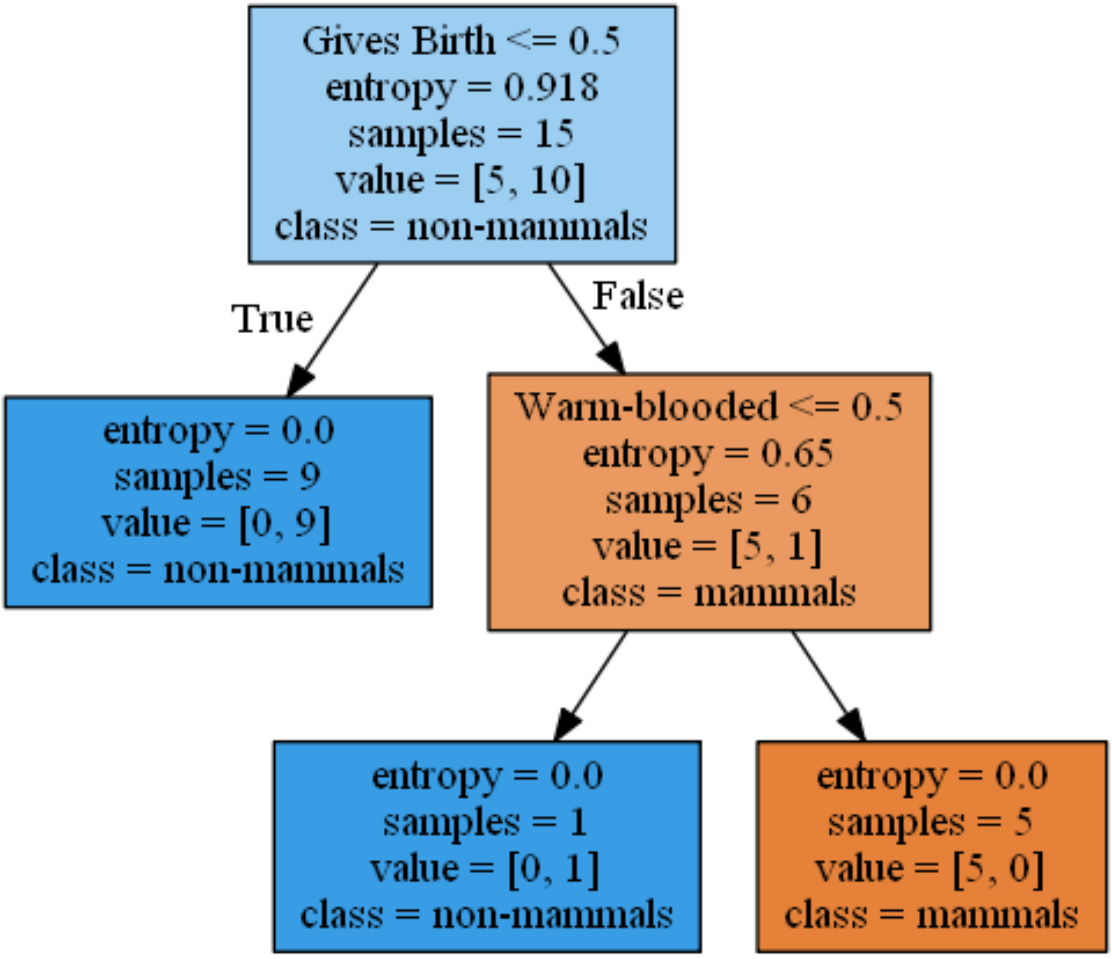
```
1            1
2            0
3            0
4            1
5            0
6            1
7            0
8            0
9            0
10           0
11           0
12           1
13           0
14           1
```

In [23]:
```
clf=tree.DecisionTreeClassifier(criterion='entropy',max_depth=3)
clf=clf.fit(x,y)
```

In [24]:
```
import pydotplus
from IPython.display import Image
```

In [25]:
```
dot_data=tree.export_graphviz(clf, feature_names=x.columns,class_names=['mammals','no
graph=pydotplus.graph_from_dot_data(dot_data)
Image(graph.create_png())
```

Out[25]:


Gives Birth <= 0.5
entropy = 0.918
samples = 15
value = [5, 10]
class = non-mammals
True
False
entropy = 0.0
samples = 9
value = [0, 9]
class = non-mammals
Warm-blooded <= 0.5
entropy = 0.65
samples = 6
value = [5, 1]
class = mammals
entropy = 0.0
samples = 1
value = [0, 1]
class = non-mammals
entropy = 0.0
samples = 5
value = [5, 0]
class = mammals

In [ ]: