decision tree for titanic

May 11, 2023

[4]:
```
import pandas as pd
df=pd.read_csv(r'C:\Users\Rakesh\Downloads\titanic.csv')
df.head()
```

[4]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th…	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

[17]:
```
x=df.
↪drop(columns=['Survived','PassengerId','Name','SibSp','Parch','Ticket','Cabin','Embarked'])
y=df['Survived']
```

[18]:
```
x.head()
```

[18]:

	Pclass	Sex	Age	Fare
0	3	male	22.0	7.2500
1	1	female	38.0	71.2833
2	3	female	26.0	7.9250
3	1	female	35.0	53.1000
4	3	male	35.0	8.0500

[19]:
```
x.describe()
```

[19]:
```
            Pclass         Age        Fare
count  891.000000  714.000000  891.000000
mean     2.308642   29.699118   32.204208
std      0.836071   14.526497   49.693429
min      1.000000    0.420000    0.000000
25%      2.000000   20.125000    7.910400
50%      3.000000   28.000000   14.454200
75%      3.000000   38.000000   31.000000
max      3.000000   80.000000  512.329200
```

[20]:
```
x.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 4 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Pclass  891 non-null    int64
 1   Sex     891 non-null    object
 2   Age     714 non-null    float64
 3   Fare    891 non-null    float64
dtypes: float64(2), int64(1), object(1)
memory usage: 28.0+ KB
```

[21]:
```
df.isnull().sum()
```

[21]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

[34]:
```
x.Age=x.Age.fillna(x.Age.mean())
```

[43]:
```
x.isnull().sum()
```

[43]:
```
Pclass    0
Sex       0
```

```
Age       0
Fare      0
dtype: int64
```

[35]:
```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
```

[36]:
```
x['Pclass']=le.fit_transform(x.Pclass)
x['Sex']=le.fit_transform(x.Sex)
x['Age']=le.fit_transform(x.Age)
x['Fare']=le.fit_transform(x.Fare)
x.head()
```

[36]:

	Pclass	Sex	Age	Fare
0	2	1	28	18
1	0	0	51	207
2	2	0	34	41
3	0	0	47	189
4	2	1	47	43

[37]:
```
from sklearn.tree import DecisionTreeClassifier
model=DecisionTreeClassifier()
```

[38]:
```
model.fit(x,y)
```

[38]:
```
DecisionTreeClassifier()
```

[39]:
```
model.score(x,y)
```

[39]:
```
0.9797979797979798
```

[40]:
```
model.predict([[0,0,28,207]])
```

```
C:\Users\Rakesh\anaconda3\lib\site-packages\sklearn\base.py:450: UserWarning: X
does not have valid feature names, but DecisionTreeClassifier was fitted with
feature names
  warnings.warn(
```

[40]:
```
array([1], dtype=int64)
```

[ ]: