Branch: master ▾

Find file | Copy path

Class / DATA602NaiveBayes Tennis Weather Prediction.ipynb

tonydiana1 Add files via upload

6aeea6f on Sep 2, 2019

1 contributor

Raw | Blame | History

257 lines (257 sloc) 7.89 KB

In [4]:
```
import pandas as pd
data = pd.read_csv("f:/machine learning/tennis.csv")
data.info()
data.columns
data.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14 entries, 0 to 13
Data columns (total 5 columns):
outlook     14 non-null object
temp        14 non-null object
humidity    14 non-null object
windy       14 non-null bool
play        14 non-null object
dtypes: bool(1), object(4)
memory usage: 542.0+ bytes
```

Out[4]:

	outlook	temp	humidity	windy	play
0	sunny	hot	high	False	no
1	sunny	hot	high	True	no
2	overcast	hot	high	False	yes
3	rainy	mild	high	False	yes
4	rainy	cool	normal	False	yes

In [5]:
```
# outlook_count = data.groupby(['outlook', 'play']).size()
# outlook_total = data.groupby(['outlook']).size()
# temp_count = data.groupby(['temp', 'play']).size()
# temp_total = data.groupby(['temp']).size()
# humidity_count = data.groupby(['humidity', 'play']).size()
# humidity_total = data.groupby(['outlook']).size()
# windy_count = data.groupby(['windy', 'play']).size()
# windy_total = data.groupby(['windy']).size()
# print(outlook_count)
# print(windy_total)
# print(outlook_total)
# print(temp_count)
# print(temp_total)
# print(humidity_count)
# print(humidity_total)
# print(windy_count)
# print(windy_total)
# p_over_yes = outlook_count['overcast','yes']
# p_over_no = 0
# p_rainy_yes = outlook_count['rainy','yes']
# p_rainy_no = outlook_count['rainy','no']
# p_rainy_yes = outlook_count['sunny', 'yes']
X_train = pd.get_dummies(data[['outlook', 'temp', 'humidity', 'windy']])
y_train = pd.DataFrame(data['play'])

#assigning predictor and target variables
#x= np.array([[-3,7],[1,5], [1,2], [-2,0], [2,3], [-4,0], [-1,1], [1,1], [-2,2], [2,7],
 [-4,1], [-2,7]])
#Y = np.array([3, 3, 3, 3, 4, 3, 3, 4, 3, 4, 4, 4])
print(X_train.info())
print(X_train.head())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14 entries, 0 to 13
Data columns (total 9 columns):
windy               14 non-null bool
outlook_overcast    14 non-null uint8
outlook_rainy       14 non-null uint8
outlook_sunny       14 non-null uint8
temp_cool           14 non-null uint8
temp_hot            14 non-null uint8
temp_mild           14 non-null uint8
humidity_high       14 non-null uint8
humidity_normal     14 non-null uint8
dtypes: bool(1), uint8(8)
memory usage: 206.0 bytes
None
   windy  outlook_overcast  outlook_rainy  outlook_sunny  temp_cool  temp_hot  \
0  False                 0              0              1          0         1
1   True                 0              0              1          0         1
2  False                 1              0              0          0         1
3  False                 0              1              0          0         0
4  False                 0              1              0          1         0

   temp_mild  humidity_high  humidity_normal
0          0              1                0
1          0              1                0
2          0              1                0
3          1              1                0
4          0              0                1
```

In [6]:
```
#Import Library of Gaussian Naive Bayes model
from sklearn.naive_bayes import GaussianNB
from sklearn.naive_bayes import GaussianNB
import numpy as np
#Create a Gaussian Classifier
model = GaussianNB()

# Train the model using the training sets
model.fit(X_train, y_train)

#Predict Output
predicted= model.predict([[False,1,0,0,0,1,0,1,0]])
print (predicted)
```

```
['yes']
```

```
C:\Users\Owner\Anaconda3\lib\site-packages\sklearn\utils\validation.py:761: DataConversio
nWarning: A column-vector y was passed when a 1d array was expected. Please change the sh
ape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
```

```
In [ ]:
```