

Name : Apurv Sarode
UID : 2019130054
Subject : AIML
Experiment : 4

Aim : For a given problem statement classify using Naïve Bayes Algorithm.

Theory:

Naïve Bayes

- Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.
- It is mainly used in text classification that includes a high-dimensional training dataset.
- Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.
- It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.
- Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles.

The Naïve Bayes algorithm is comprised of two words Naïve and Bayes, Which can be described as:

Naïve: It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is identified on the bases of colour, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other.

Bayes: It is called Bayes because it depends on the principle of Bayes' Theorem.

Program:

```
import numpy as np
import pandas as pd
```

```
data = pd.read_csv('play_tennis.csv')
data.head
data.drop(columns=['day'],inplace=True)
```

data

	outlook	temp	humidity	wind	play
0	Sunny	Hot	High	Weak	No
1	Sunny	Hot	High	Strong	No
2	Overcast	Hot	High	Weak	Yes
3	Rain	Mild	High	Weak	Yes
4	Rain	Cool	Normal	Weak	Yes
5	Rain	Cool	Normal	Strong	No
6	Overcast	Cool	Normal	Strong	Yes
7	Sunny	Mild	High	Weak	No
8	Sunny	Cool	Normal	Weak	Yes
9	Rain	Mild	Normal	Weak	Yes
10	Sunny	Mild	Normal	Strong	Yes
11	Overcast	Mild	High	Strong	Yes
12	Overcast	Hot	Normal	Weak	Yes
13	Rain	Mild	High	Strong	No

```
data['play'].value_counts()
```

```
Yes    9  
No     5  
Name: play, dtype: int64
```

```
p_y=9/14  
p_n=5/14
```

```
print(p_y)  
print(p_n)
```

```
0.6428571428571429  
0.35714285714285715
```

```
pd.crosstab(data['outlook'],data['play'])
```

	play	No	Yes
outlook			
Overcast	0	4	
Rain	2	3	
Sunny	3	2	

```
p_overcast_n=0  
p_rain_n=2/5  
p_sunny_n=3/5
```

```
p_overcast_y=4/9  
p_rain_y=3/9  
p_sunny_y=2/9
```

```
pd.crosstab(data['temp'],data['play'])
```

	play	No	Yes
temp			
Cool	1	3	
Hot	2	2	
Mild	2	4	

```
p_cool_n=1/5  
p_hot_n=2/5  
p_mild_n=2/5
```

```
p_cool_y=3/9  
p_hot_y=2/9  
p_mild_y=4/9
```

```
pd.crosstab(data['humidity'],data['play'])
```

	play	No	Yes
humidity			
High	4	3	
Normal	1	6	

```
p_high_n=4/5
p_normal_n=1/5

p_high_y=3/9
p_normal_y=6/9
```

```
pd.crosstab(data['wind'],data['play'])
```

play	No	Yes
wind		
Strong	3	3
Weak	2	6

```
p_strong_n=3/5
p_weak_n=2/5

p_strong_y=3/9
p_weak_y=6/9
```

```
Outlook=input('Is the outlook is overcast or rain ot sunny')
```

```
if Outlook=='overcast':
    A_y=p_overcast_y
    A_n=p_overcast_n
```

```
elif Outlook=='rain':
    A_y=p_rain_y
    A_n=p_rain_n
```

```
elif Outlook=='sunny':
    A_y=p_sunny_y
    A_n=p_sunny_n
```

```
else:
    print('Enter valid outlook')

print('The probability of',Outlook,'is',A_y)
```

The probability of overcast is 0.4444444444444444

```
Tempreture=input('Is the tempreture is cool or hot ot mild')
```

```
if Tempreture=='cool':
    B_y=p_cool_y
    B_n=p_cool_n
```

```
elif Tempreture=='hot':
    B_y=p_hot_y
    B_n=p_hot_n
```

```
elif Tempreture=='mild':
    B_y=p_mild_y
    B_n=p_mild_n
```

```
else:
    print('Enter valid temp')

print('The probability of',Tempreture,'is',B_y)
```

The probability of mild is 0.4444444444444444

```

humidity=input('Is the humidity high or normal')

if humidity=='high':
    C_y=p_high_y
    C_n=p_high_n

elif humidity=='normal':
    C_y=p_normal_y
    C_n=p_normal_n

else:
    print('Enter valid humidity')

print('The probability of',humidity,'is',C_y)

```

The probability of high is 0.3333333333333333

```

wind=input('Is the wind strong or weak')

if wind=='strong':
    D_y=p_strong_y
    D_n=p_strong_n

elif wind=='weak':
    D_y=p_weak_y
    D_n=p_weak_n

else:
    print('Enter valid wind state')

print('The probability of',wind,'is',D_y)

```

The probability of strong is 0.3333333333333333

```

P_yes=p_y*A_y*B_y*C_y*D_y
print('probability of playing tennis today is',P_yes)

P_no=p_n*A_n*B_n*C_n*D_n
print('probability of not playing tennis today is',P_no)

if P_yes>P_no:
    print('You can play tennis!')

else:
    print("You can't play tennis")

```

probability of playing tennis today is 0.014109347442680775
probability of not playing tennis today is 0.0
You can play tennis today!

Conclusion:

In the above experiment of AIML Lab, I learnt about Naïve Bayes Algorithm. This algorithm is used for supervised learning models. This algorithm is easy to use as we make assumption that the attributes are independent and then classify the input.