

Experiment-4

Samkit Shah
2019130060
TE Comps
Batch C

Aim:

To train and test machine learning models using naive bayes algorithm.

Theory:

- The Bayes' Theorem is used to create a collection of classification algorithms known as Naive Bayes classifiers. It is a family of algorithms that share a similar idea, namely that each pair of features being classified is independent of the others.
- The Naive Bayes assumption is that each feature contributes equally and independently to the outcome.
- The Bayes' Theorem calculates the likelihood of an event occurring given the probability of a previous event.

Code:

```
import pandas as pd
data = pd.read_csv('tennis.csv')
data
outlook = data.groupby(['outlook', 'play']).size()
temp = data.groupby(['temp', 'play']).size()
humidity = data.groupby(['humidity', 'play']).size()
windy = data.groupby(['windy', 'play']).size()
play = data.play.value_counts()
print(temp)
print("\n")
print(humidity)
print("\n")
print(windy)
print("\n")
print(outlook)
print("\n")
print('play')
print(play)
total_y = play["yes"]
total_n = play["no"]

play_total = total_y + total_n
def find_prob(outlook_val, temp_val, humidity_val, windy_val, play_val):
    p_outlook_play = outlook[outlook_val][play_val]/play[play_val]
    p_temp_play = temp[temp_val][play_val]/play[play_val]
    p_humidity_play = humidity[humidity_val][play_val]/play[play_val]
    p_windy_play = windy[windy_val][play_val]/play[play_val]
    p_play = play[play_val]/play_total

    prob = p_outlook_play * p_temp_play * p_humidity_play * p_windy_play * p_play
```

```

    return prob
def pred_play(outlook_val, temp_val, humidity_val, windy_val):
    prob_yes = find_prob(outlook_val, temp_val, humidity_val, windy_val, "yes")
    prob_no = find_prob(outlook_val, temp_val, humidity_val, windy_val, "no")

    print("Probability that you will play Tennis: ", prob_yes/(prob_yes+prob_no))
    print("Probability that you won't play Tennis: ", prob_no/(prob_yes+prob_no))

    if prob_yes > prob_no:
        print("You should play Tennis today!")

    else:
        print("You should not play Tennis today!")
outlook_value = 'sunny' #"overcast", "sunny", "rainy"
temp_value = 'mild' #"cool", "hot", "mild"
humidity_value = 'normal' #"high", "normal"
windy_value = False #"False", "True"
pred_play(outlook_value, temp_value, humidity_value, windy_value)

```

Output:

```

Probability that you will play Tennis: 0.8045052292839904 Probability that you
won't play Tennis: 0.1954947707160097 You should play Tennis today!

```

Conclusion:

- I learned about the basic Bayes theorem through the naive bayes experiment above. The likelihood of an event occurring in relation to any condition is described by Bayes' theorem. In the naive bayes method, we calculate the probability of each output category and choose the one with the highest probability.
- The naive bayes technique is based on two assumptions: each data point in the dataset adds to the dataset independently and equally.