Lab 4 Set 1 – (Bayes theorem and Bayesian Classifier)

Objective: Finding probability by applying Bayes theorem and applying Bayesian Classifier on dataset for finding probability of an event and accuracy of model.

Question-1: (20 Minutes) **1 Mark**

A patient goes to see a doctor. The doctor performs a test with 99 percent reliability--that is, 99 percent of people who are sick test positive and 99 percent of the healthy people test negative. The doctor knows that only 1 percent of the people in the country are sick.

(a) If the patient tests positive, what are the chances the patient is sick?

The intuitive answer is 99 percent, but the correct answer is 50 percent...."

posterior = (likelihood * prior) / evidence

Considering the hypothetical probablity of	10000 people			
	Diseased	**Not Diseased**		
Test +	99	99	198	P(B)
Test -	1	9,801	9,802	
	100	9,900	10,000	
	P(A)			
A - event of disease	P(A)	100/10000	0.01	
B - event of +ve test	P(B)	198/10000	0.0198	
B/A - event of observing event +ve if diseased is true	P(B/A)=P(A^B)/P(A)	(99/10000)*(0.01)	0.99	
A/B - event of observing event diseased if +ve is true	P(A/B)			

P(A | B) = (0.99 x 0.01) / 0.0198 = 0.50 = 50%.

```
1   TESTpos =  99
2   TESTneg = 99
3   TESTndandneg = 1
4   totalcase = 10000
5   #totalp = 100
6   TESTndandneg = totalcase - (TESTpos + TESTneg+TESTndandneg)
7   A = TESTpos/totalcase
8   B = (TESTpos+TESTneg)/totalcase
9   prob_b_a = (TESTpos/totalcase)/A
10  prob_a_b = (prob_b_a*A)/B
11  print(prob_a_b*100,"%")
```

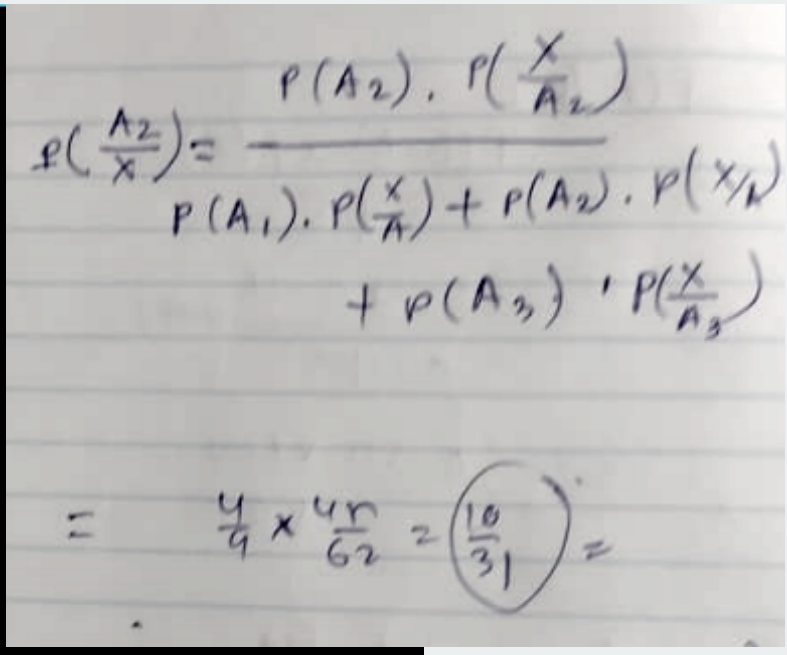
```
50.0 %
```

Question-2: (20 Minutes) **1 Mark**

Three identical boxes contain red and white balls. The first box contains 3 red and 2 white balls, the second box has 4 red and 5 white balls, and the third box has 2 red and 4 white balls. A box is chosen very randomly and a ball is drawn from it. If the ball that is drawn out is red.

(a) What will be the probability that the second box is chosen?

```
#fist box
a = 3 # 3red balls
b = 2 # 2white balls
#second box
c = 4 # 4red balls
d = 5 # 5white balls
#third box
e = 2 # 2red balls
f = 4 # 4white balls
b1 = 1/3
b2 = 1/3
b3 = 1/3
# red/total
x_b1 = a/(a+b)
x_b2 = c/(c+d)
x_b3 = e/(e+f)
s = b2*x_b2
ss = (b1*(x_b1))+(b2*(x_b2))+(b3*(x_b3))
pa2_x = s/ss
print(pa2_x)
```




```
0.32258064516129037
```

Question-1: Write programs for followings questions with the help of below tabular data(save this tabular data with "Position_Salaries.csv" name in your system.

(a) Draw line of best fit using Polynomial Regression approach/classifier using 2 and 3 degrees.

(b) Find RMSE and R2 score

Data

	Position	Level	Salary
0	Business Analyst	1	45000
1	Junior Consultant	2	50000
2	Senior Consultant	3	60000
3	Manager	4	80000
4	Country Manager	5	110000
5	Region Manager	6	150000
6	Partner	7	200000
7	Senior Partner	8	300000
8	C-level	9	500000
9	CEO	10	1000000

Position	Level	Salary
Business Analyst	1	45000
Junior Consultant	2	50000
Senior Consultant	3	60000
Manager	4	80000
Country Manager	5	110000
Region Manager	6	150000
Partner	7	200000
Senior Partner	8	300000
C-level	9	500000
CEO	10	1000000

additional.py | Position_Salaries.csv

Position_Salaries.csv

```
1  Position,Level,Salary
2  Business Analyst,1,45000
3  Junior Consultant,2,50000
4  Senior Consultant,3,60000
5  Manager,4,80000
6  Country Manager,5,110000
7  Region Manager,6,150000
8  Partner,7,200000
9  Senior Partner,8,300000
10 C-level,9,500000
11 CEO,10,1000000
```

additional.py | Position_Salaries.csv

additional.py > ...

```
1  from sklearn.metrics import mean_squared_error
2  from math import sqrt
3  from sklearn.metrics import r2_score
4  import pandas as pd
5  import matplotlib.pyplot as plt
6  import numpy as np
7  a = pd.read_csv(r"Position_Salaries.csv")
8  print(a)
9  s = a["Salary"]
10 ss = a["Level"]
11 plt.scatter(s,ss,color="black")
12 plt.plot(s,ss,color="c")
13 s2 = np.array(s)
14 sw = np.array(ss)
15 c , d = np.polyfit(s2,sw,1)
16 plt.plot(s2,s2*c+d,color = 'black')
17 plt.show()
18 m = sqrt(mean_squared_error(s2, sw))
19 print("RMSE:",m)
20 m2 = r2_score(s2,sw)
21 print("R2 SCORE:",m2)
```