IC-252 Lab

Lab Assignment-3

due for submission on Moodle by 12th March

Question 1: The given dataset 'data.csv' consists of some weather data over the past 100 days for a region in Himachal Pradesh. It consists of columns - Day, Cloudy, Rain and Snow. The column 'Day' has days numbered from 1-100. If a given day was cloudy it is denoted by '1' else '0'. If it rained greater than 0.01 inches on a particular day, it is represented as '1' else '0'. Similarly, if it snows, it is represented by '1' else '0'.

With the given information, calculate the following:

A. Probability that a day is: (i) cloudy (ii) raining (iii) snowing
B. Probability that it will rain given that it is a cloudy day and also the Probability that it is cloudy given it is raining. Using the values obtained, verify the formula:
   P(A|B) = P(A) P(B|A)P(B)
C. Probability of a sun shower i.e. it is raining given it is not cloudy.
D. Probability that it will either rain or snow, given it is a cloudy day.
E. Probability that it will both rain and snow, given it is a cloudy day.

Question 2: Suppose that a laboratory test to detect a certain disease has the following statistics.
A = event that the person has the disease
B = event that the test result is positive
It is known that P (B|A) = 0.95 and $P(B|\bar{A})$ = 0.05, and 1% of the population actually has the disease.
(Note: The question is similar q1 of the previous assignment with minor modifications)
a) Theoretically calculate the probability that a person tested positive actually has the disease
b) Simulate the experiment in python.
Take the size of Population(N) as 1000, 10000, 100000.
Calculate the probability(P) experimentally for the given values of N.
c) Plot the Graph of P v/s N (domain of N should from 1 to 100000)
[Hint: Make two lists, one of population(healthy/diseased) and second of test_results(positive/negative).
Use random.random() for sampling]

Question 3: Let
H = event that person has disease,
S = event that test result is positive
R1 = event that positive test result came from Machine 1
R2 = event that positive test result came from Machine 2

P(H)=0.001, P(R1)=0.2, P(R2)=0.8,
P(R1|H)=0.89, P(R2|H)=0.99, P(R1|$\bar{H}$)=0.025, P(R2|$\bar{H}$)=0.005

Given that person has a disease and test result is positive. What is the probability that the result came from Machine 2? Simulate it 10 times and see the difference in probabilities.

Question 4:

We believe there are three types of managers:
underperformers, in-line performers, and outperformers. The underperformers (MU) beat the market only 25% of the time, the in-line performers (MI) beat the market 50% of the time, and the outperformers (MO) beat the market 75% of the time.

Initially we believe a given manager is most likely to be an in-line performer, and is less likely to be an underperformer or an outperformer. Specifically, our prior belief is that a manager has a 60% chance of being an in-line performer, a 20% chance of being an underperformer, and a 20% chance of being an outperformer. We can summarize this as:

P[MU]=P[p=0.25]=20%, P[MI]=P[p=0.50]=60%, P[MO]=P[p=0.75]=20% ….(1)

a.) By taking the values given in (1), suppose the manager beats the market two years in a row. What should our updated beliefs be?

b.) After updating the beliefs from part (a), what if the manager again beats the market next year, what should be the updates now?

c.) By taking the values given in (1), suppose the manager beats the market three years in a row. What should our updated beliefs be? How is it different from part (b)?

For the lab- Simulate part (b) for the next 10 years.