ML HW1

Kaushal Vinay Nerkar (knerkar@binghamton.edu)

Question 1 (Exercise 1.7)

Exercise 1.7

(a) always returns •

agree all three points → f_8

agrees 2 points → f_7, f_6, f_4

agrees 1 point → f_5, f_3, f_2

none of them → f_1

always returns o

agree 3 points → f_1

agree 2 points → f_2, f_3, f_5

agree 1 point → f_7, f_6, f_4

~~agr~~ none of them → f_8

(b) for
~~??~~agrees all three points
$\rightarrow f_1$
2points $\rightarrow f_2, f_3, f_5$
1 point $\rightarrow f_4, f_6, f_7$

(c) agrees to all 3 points when XOR is used.
3 points $\rightarrow f_2$
2 points $\rightarrow f_1, f_4, f_6$
1 points $\rightarrow f_5, f_8, f_3$
0 points $\rightarrow f_7$

(d)
3 points $\rightarrow f_7$
2 points $\rightarrow f_5, f_8, f_3$
1 points $\rightarrow f_1, f_4, f_6$
0 points $\rightarrow f_2$

Question 2 (Exercise 1.8)

Machine learning

Exercise 1.8

given $\mu = 0.9$

and $\nu \le 0.1$ for red marble

$n = 10$

so we can have atmost 1 red marble

so $P(\text{red} \le 1) = P(0) + P(1)$

using binomial distribution $\Rightarrow {}^{n}C_{r}\, \mu^{r} (1-\mu)^{n-r}$ . . .

$$P(0) + P(1) = (1-\mu)^{10} + 10\mu(1-\mu)^{9}$$

$$= (0.1)^{10} + (0.1)^{9} \times 9$$

$$= (0.1)^{9}\,(9.1)$$

$$= 10 \times 10^{-9}$$

$$= 10^{-8} \quad \text{(approx)}$$

Question 3 (Exercise 1.9)

Excercise 1.9

given $\mu = 0.9$, $N = 10$

$\nu \leq 0.1$

$|\mu - \nu| = 0.8$

so we know $|\mu - \nu| > \varepsilon$ & $\varepsilon > 0$

so lets choose $\varepsilon = 0.7$

so $P(|\mu - \nu| > \varepsilon) \leq 2e^{-2\varepsilon^2 N}$

$= 2e^{-9.8} = 1.109 \times 10^{-4}$

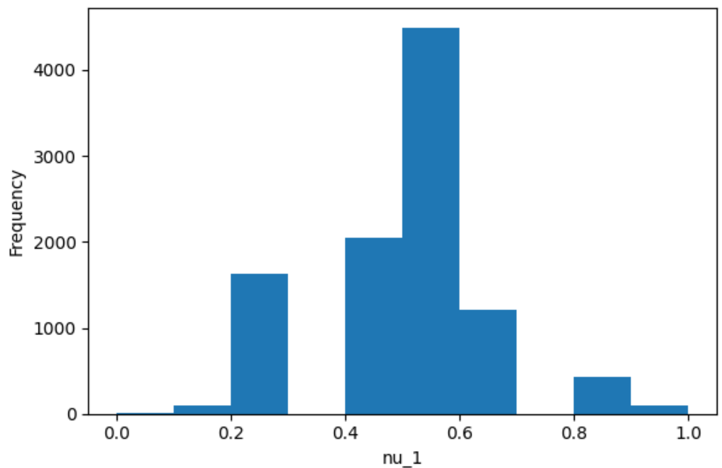
so using hoeffding Inequality we see upperbound for probability is much higher than in previous case.

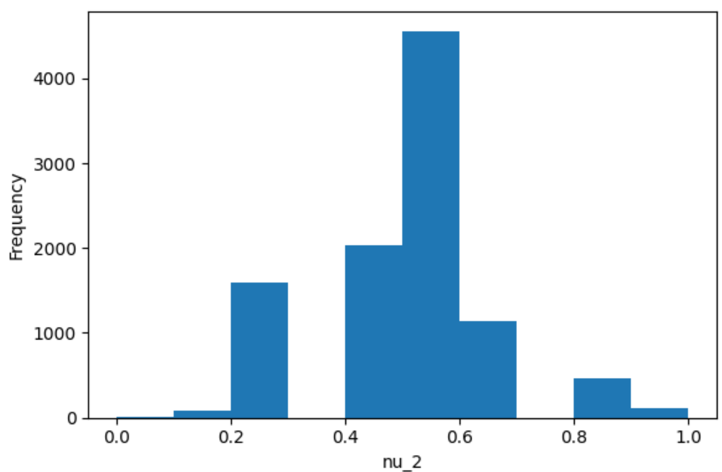
Question 4 (Exercise 1.10)

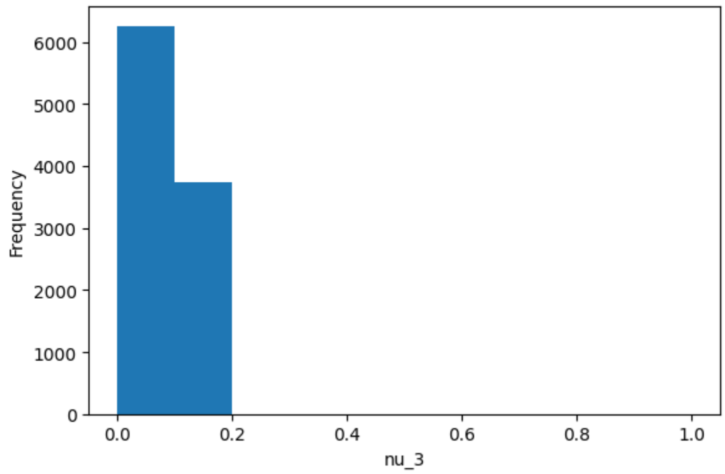
Link to notebook :
https://colab.research.google.com/drive/1eG0drpLyHQZhRyDb9_vYwTFpUOSlQLc3?usp=sharing

a. 0.5 for all coins as all are fair.

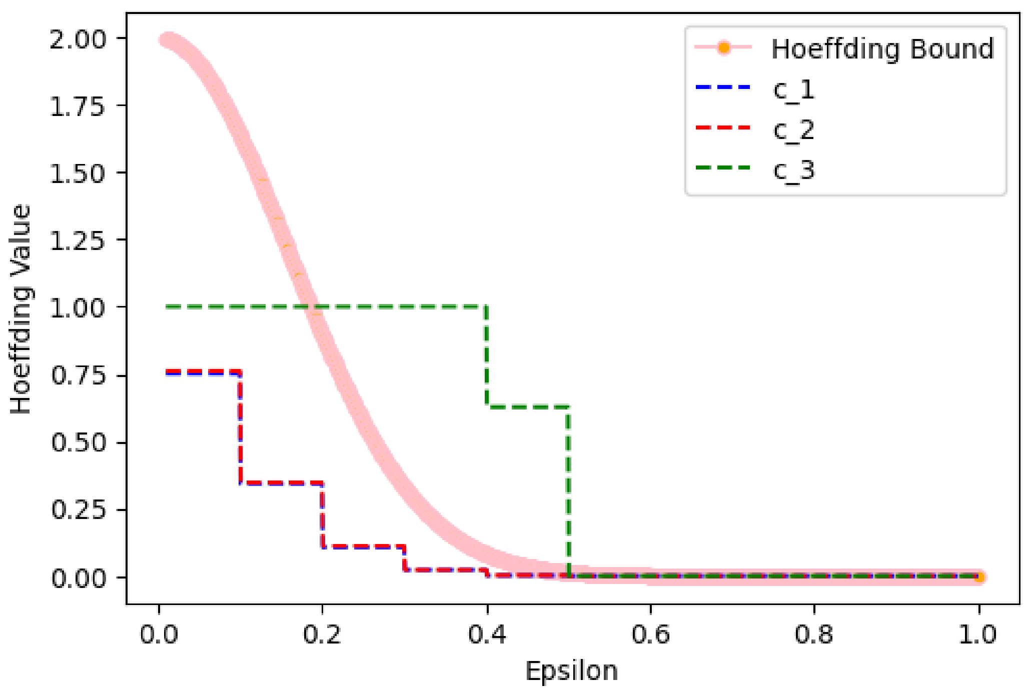
b.







c.



d. For C1 ,Crand are following hoefding's bound but Cmin(with minimum frequency of head) does not.
e. Here we are creating a similar situation like in case of multiple bins where selection is biased, or selection is done on prior knowledge or adjusting the selections. Thus, it violates the hoefding's inequality which provides a bound-on difference between actual and observed fractions of head in this case.

Question 5 (Exercise 1.12)

Excercise 1. 12

(a)

given → 4000 data points.

with this data, we can produce a hypothesis but can't gurantee the correctness of it. As real function can be very complex and cannot be determined from the given sample.

So option 'c' is corect.

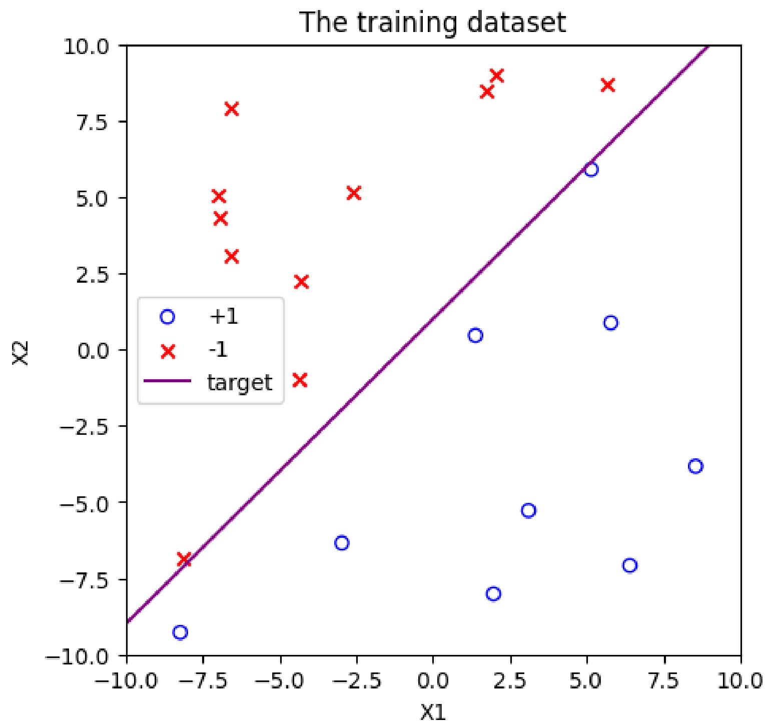
Also if we learn ~~some~~ some data using hoeffding inequality we can state that probability of f close to g is high or error in f & g is less.

We can only confirm one of 2 thing ie we can produce a hypothesis or we declare hypothesis is bad

(b)

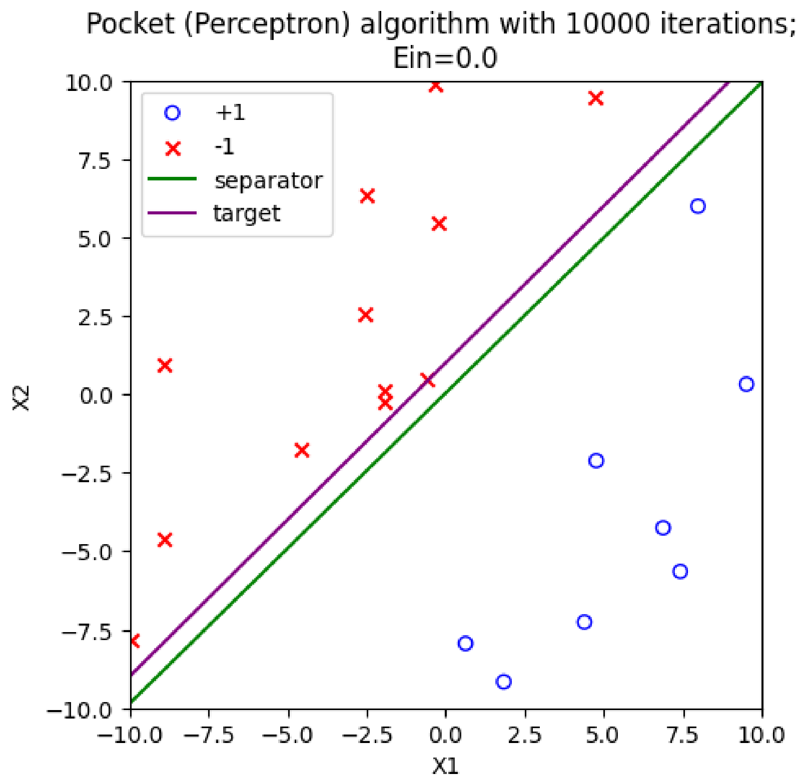
Question 6 (Problem 1.4)

a.



Link for notebook :
https://colab.research.google.com/drive/1PSjIl20-bUwo0eLW0YTwq0IP-UmwdYZN?usp=sharing

b.

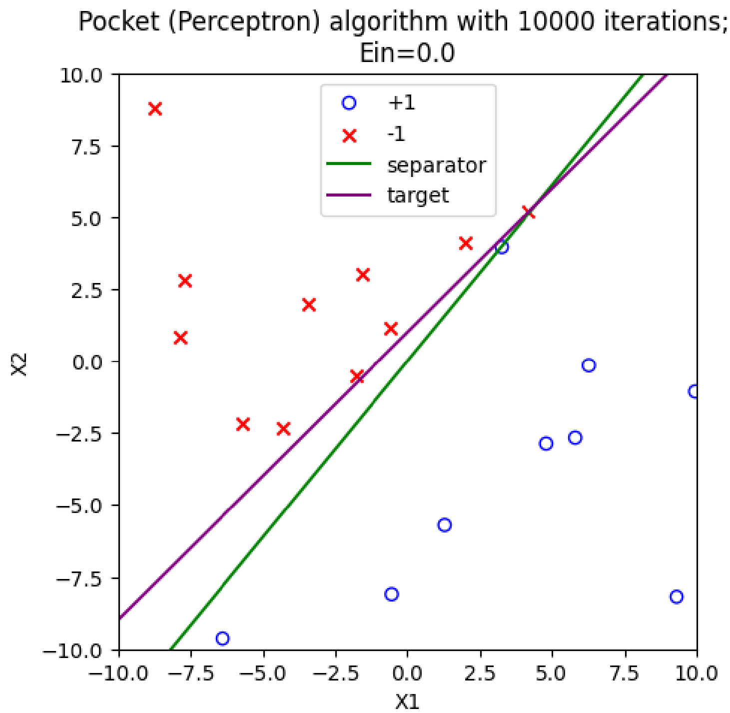



```
Ein_best 0.0,

w_best
[[  1.0000001 ]
 [ 26.47378598]
 [-26.73238471]]
Converging after  281
```

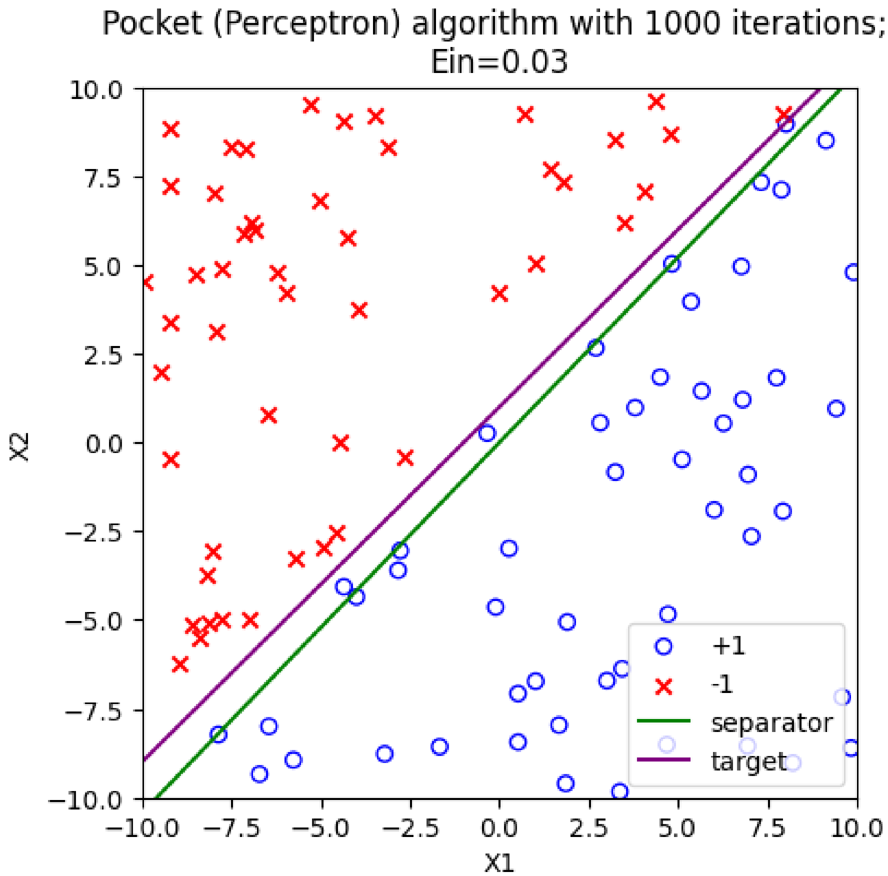
Converges after 281 iterations ,calculated iterations till ErrorIn were > 0.01.
Final Hypothesis is close to target but not the same as Final Hypothesis is also able to linearly separate the data points and algorithm won't improve once it finds such hypothesis(able to linearly separate data).

c.



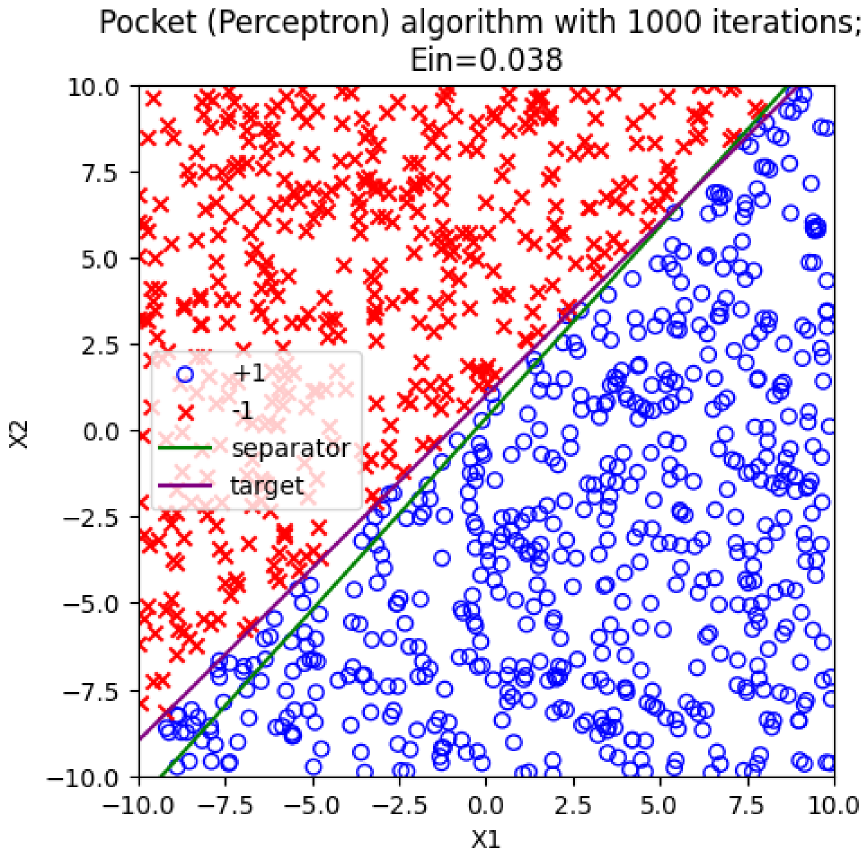
Similar results as we observed in previous case.

d.



Here it finds the Final hypothesis but iterations to converge are more than compared to previous cases.

e.



So, comparing b and e graph we see that number of iterations to find a solution increases with number of datapoints.

Question 7 (Problem 1.7)

a.

Question 1.7

Problem 1.7

$P(K|N, \mu) = {}^{N}C_{K} \; \mu^{k} (1-\mu)^{N-K}$

$\vartheta = \frac{k}{N}$

a) for $\vartheta = 0$ implies $K = 0$

So lets consider a event A where we flip coin n times to get zero heads.

So probability of getting atleast 1 heads n times would be $[1-(1-\mu)^n]$ ie $P(A) > 0$

So now we need this to be repeated for M times

$= P(A_1 = 0 + A_2 = 0 \ldots + A_M = 0)$

$= 1 - P(A_1 > 0) P(A_2 > 0) \ldots P(A_M > 0)$

$= 1 - \prod_{0}^{m} P(A_r > 0)$

$= 1 - [1-(1-\mu)^n]^M$

So for $n = 10$, $M = 1$, $y = 0.05$

$$P = 1 - \left[1 - (1 - 0.05)^{10}\right]^{1}$$

$$P = 1 - 0.401$$

$$P = 0.598$$

for $M = 1000$, $n = 10$, $y = 0.05$

$$P = 1 - \left[1 - (1 - 0.05)^{10}\right]^{1000}$$

$$P = 1 - \cancel{2}.e-397$$

$$P = 1 - 2.6 \times 10^{-397}$$

$$\approx 1$$

for $M = 1000000$ $n = 10$ $y = 0.05$

$$P = 1 - \left[1 - (0.95)^{10}\right]^{1000000}$$

$$= 1 - (0.407)^{1000000}$$

$$\approx 1$$

So for $n = 10$, $M = 1$, $\mu = 0.8$

$$P = 1 - \left[1 - (1-0.8)^{10}\right]^{1}$$
$$= 1 - \left[1 - 1.02 \times 10^{-7}\right]$$
$$= 1 - 0.999 = 1.02 \times 10^{-7}$$

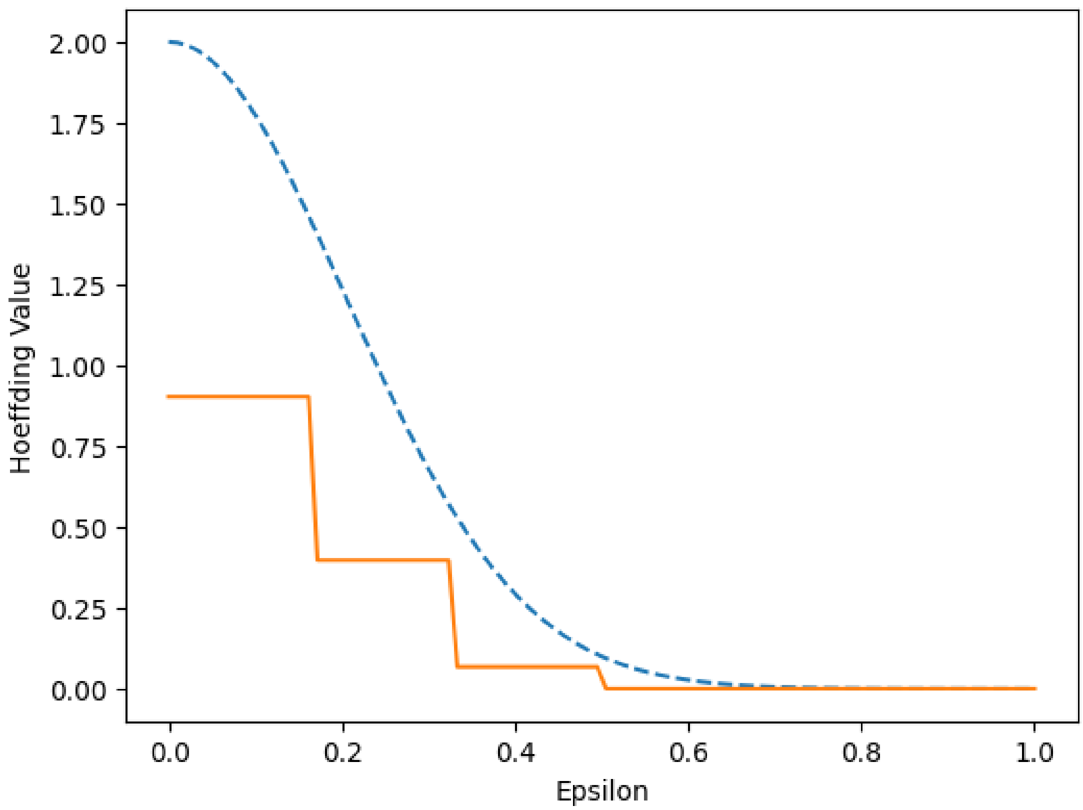
$n = 10$, $M = 1000$, $\mu = 0.8$

$$P = 1 - \left[1 - (0.2)^{10}\right]^{1000}$$
$$= 1 - 0.9998\ldots$$
$$= 0.0001$$

$n = 10$ $M = 1000\,000$ $\mu = 0.8$

$$P = 1 - 0.9026\ldots$$
$$= 0.097\ldots$$

b.



Links to Notebook

- HW Question 4 i.e. (Exercise 1.10)
  https://colab.research.google.com/drive/1eG0drpLyHQZhRyDb9_vYwTFpUOSlQLc3?usp=sharing
- HW Question 6 i.e. (Problem 1.4)
  https://colab.research.google.com/drive/1PSjIl20-bUwo0eLW0YTwq0IP-UmwdYZN?usp=sharing