Machine Learning

Assignment 1

Name: **Qazi Danish Ayub**

Roll no.: **MSDS20075**

Q1:

(a)

```
from scipy.stats import binom

prob = []
n,p ,k = 10, 0.05 , 10

coins = [1, 1000, 1000000]
for m in coins:
  ans = binom.pmf(k,n,p)
  print("For m :"+ str(m) +"  Pobability ="+ str(ans*m))
```

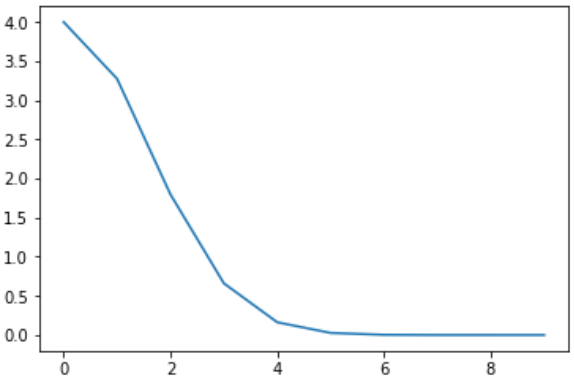
```
For m :1  Pobability =9.765625000000022e-14
For m :1000  Pobability =9.765625000000022e-11
For m :1000000  Pobability =9.765625000000022e-08
```

(b) Probability for Hoeffding inequality bound for Epsilon in range (0,1)

```
def hoeffding (x, epsilon):
    y = 4.0 * math.exp (-2*epsilon*epsilon*x)
    return y
```

```
y1=[]
y=np.arange (0, 1, 0.1)
for i in y:
    y1.append(hoeffding(10,i))
```

```
plt.plot(y1)
plt.show()
```

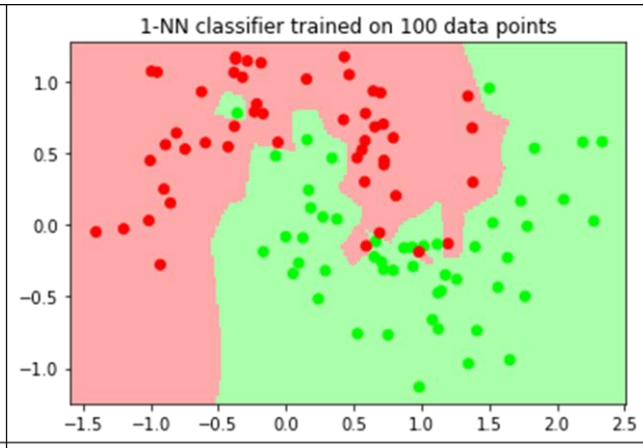
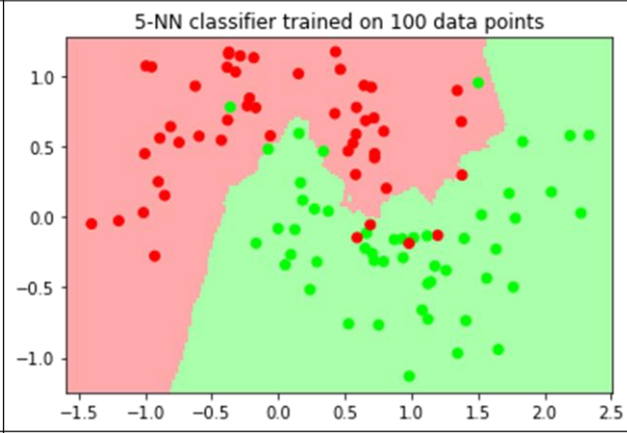
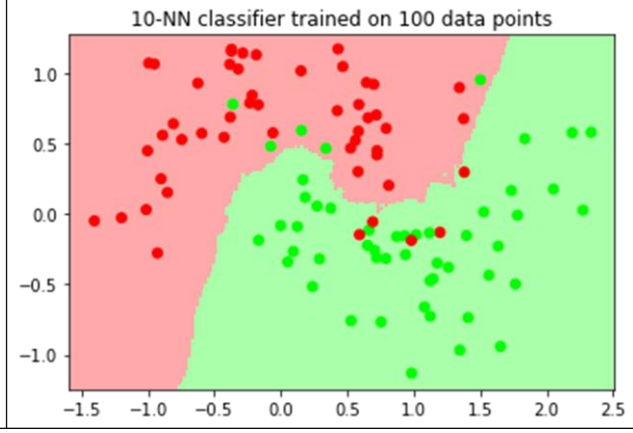


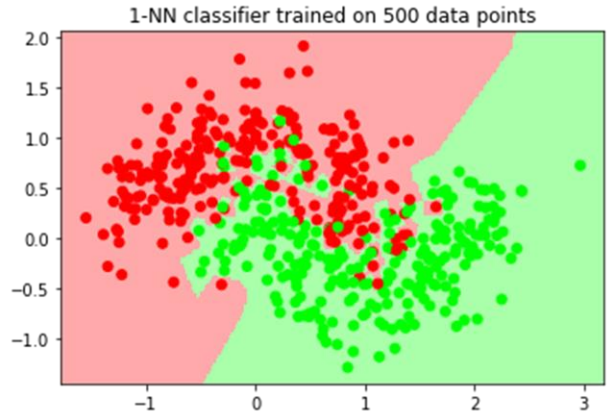
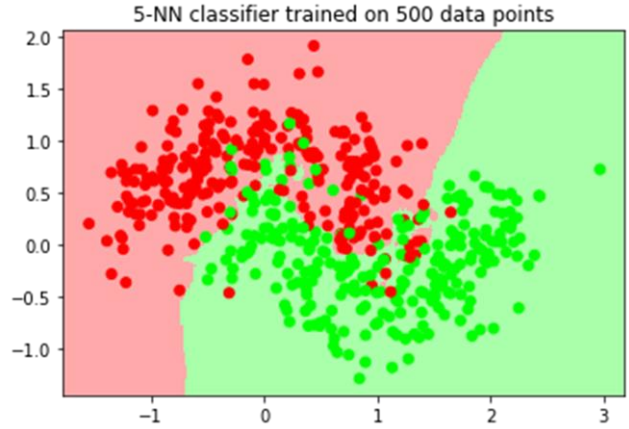
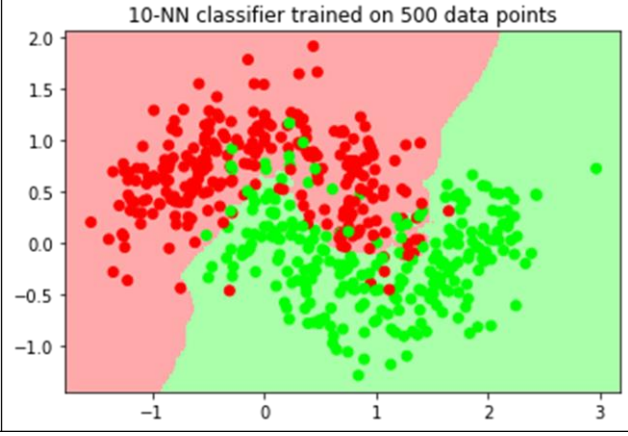
Q3.

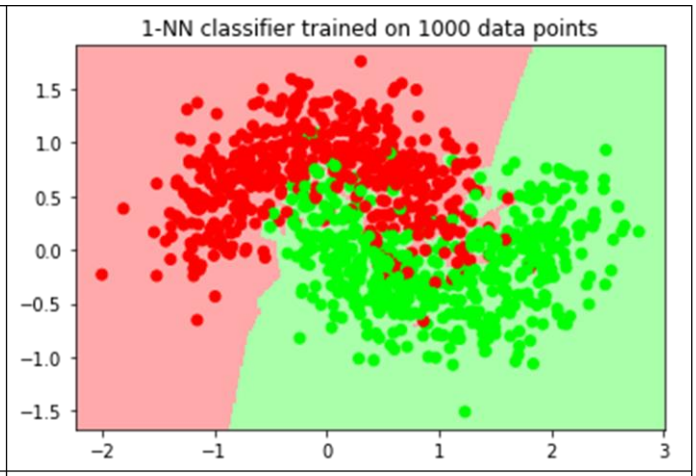
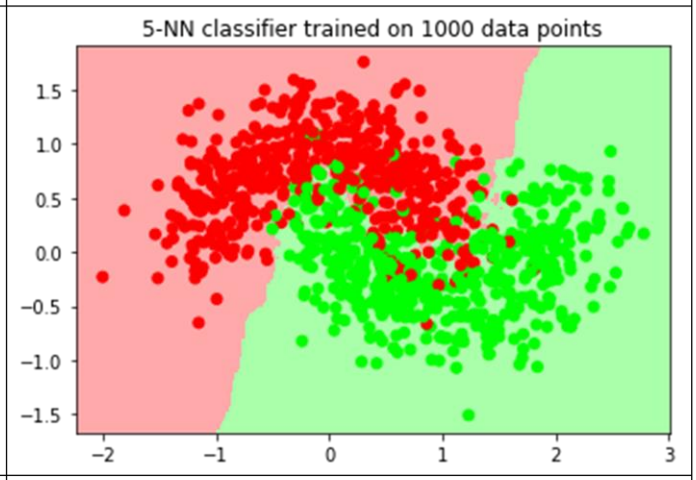
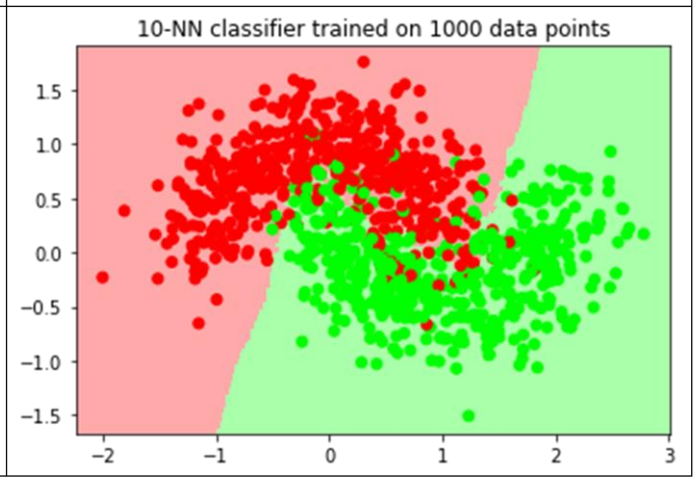
KNN (k nearest neighbor) is a classification algorithm.

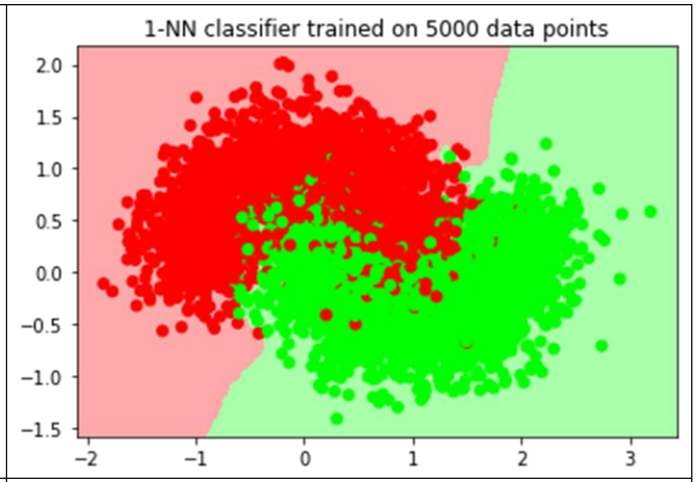
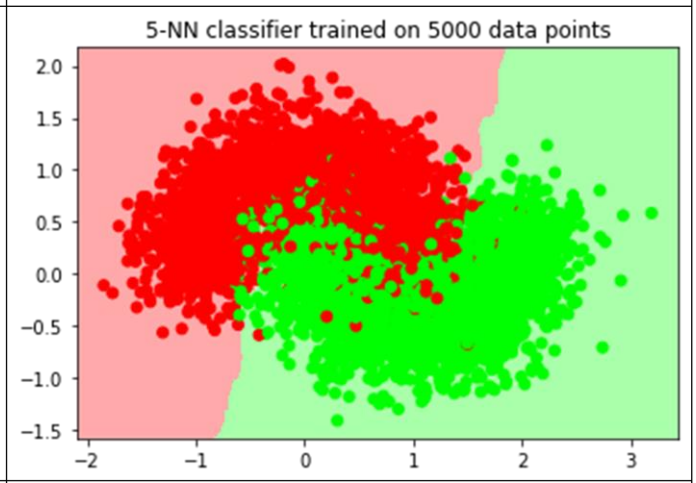
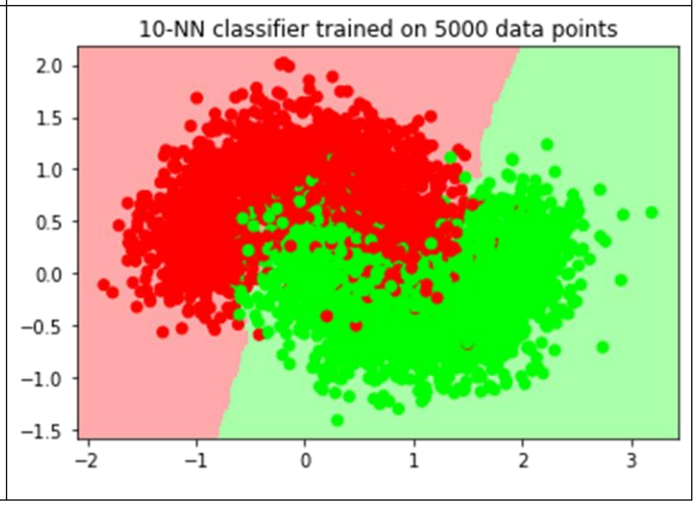
Larger values of K are less sensitive to noise and makes smoother between the boundaries of classes, by increasing K the boundary become smoother, and by increasing N number of datapoint increases. Visually Optimal values for K is in between 3 and 7.

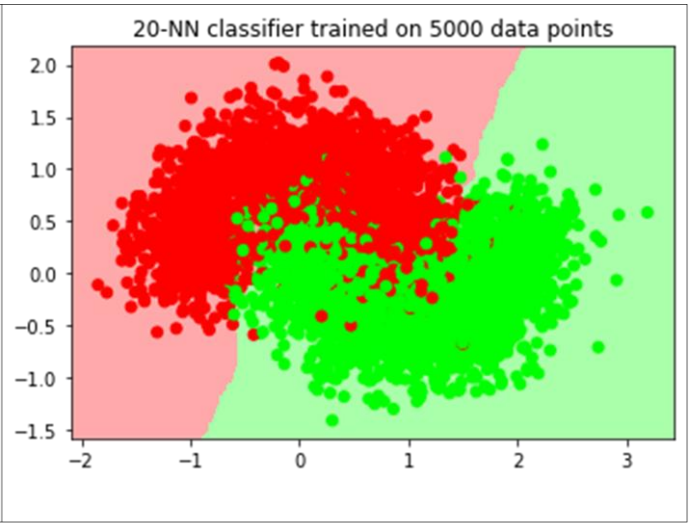
Following table shows the result of various values tested for various values of K and N

1	K=1	N=100	
2	K=5	N=100	
3	K=10	N=100	

4	K=1	N=500	
5	K=5	N=500	
6	K=10	N=500	

7	K=1	N=1000	
8	K=5	N=1000	
9	K=10	N=1000	

10	K=1	N=5000	
11	K=5	N=5000	
12	K=10	N=5000	

13	K=20	N=5000	

20-NN classifier trained on 5000 data points
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2
−1
0
1
2
3