Part I

Question 1:

a) P(a,b,c) = 0.8 * 0.8 * 0.8 = 0.512
P(a,b) = 0.8 * 0.8 = 0.64
The probability that this group makes the right decision based on majority voting is 0.64.

b) F = $\prod_{i=1}^{n} pi$ where n = (total number/ 2) + 1

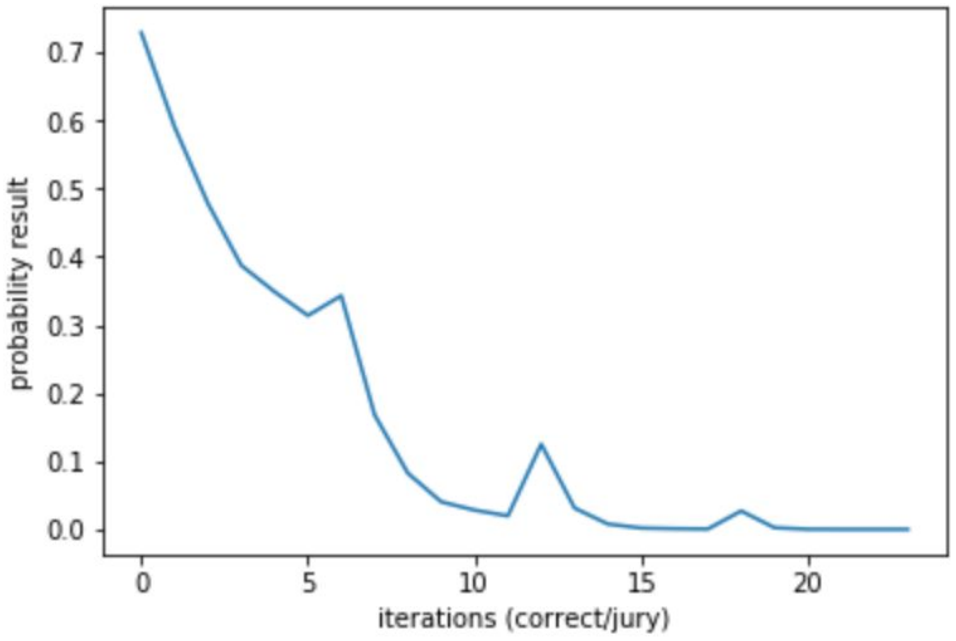
n= (21/2)+1= 10.5 + 1 =11.5 ≃ 11

F = 0.6 * 0.6 * 0.6 * 0.6 * 0.6 * 0.6 * 0.6 * 0.6 * 0.6 * 0.6 * 0.6 = 0.0036

c)

```
In [4]: m = 21

In [5]: n = int(m/2) +1

In [7]: p = 0.6

In [10]: F = 1

In [11]: for i in range(0, n):
             F = F * p

In [12]: F
Out[12]: 0.0036279705599999994
```

d) The graph below has the jury size n = [4, 8, 12, 16, 19, 21] and the probability p = [0.9, 0.7, 0.5, 0.3].



e) The highest chance to make the correct decision has the radiologist as the probability is 0.85 while for the doctors with majority voting is 0.64 and the lowest chance comes from students with 0.0036.

To make the probability of the doctors equal with the students we should have just one student as the result of doctors is 0.64 or 0.512 and one student has the probability of 0.6.

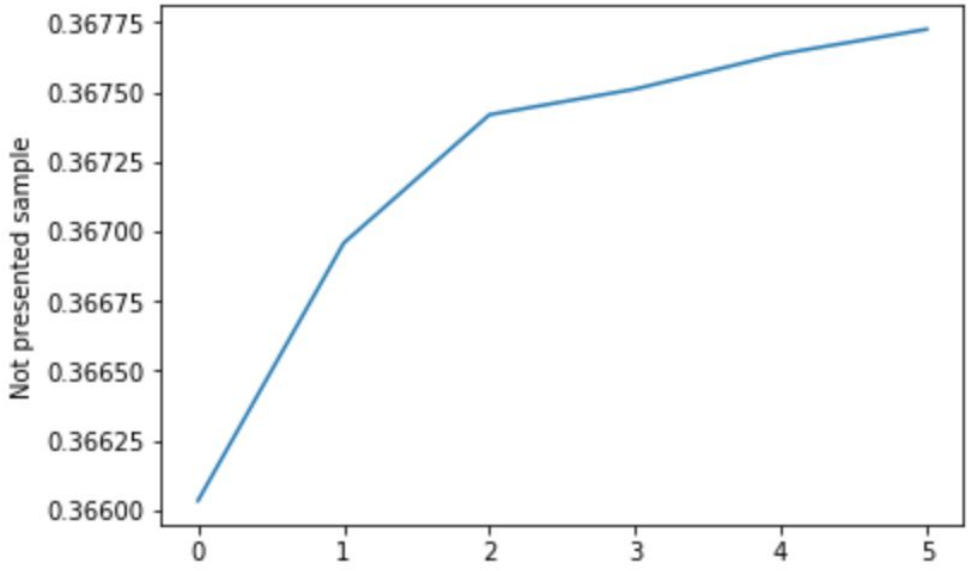
Question 2:

p1(w|x)		p2(w|x)		p3(w|x)		Mean		Max		Min		Prod	
A	B	A	B	A	B	A	B	A	B	A	B	A	B
0.9	0.1	0.9	0.1	0.0	1.0	0.6	0.4	0.9	1.0	0.0	0.1	0.0	0.01
0.9	0.1	0.9	0.1	0.3	0.7	0.7	0.3	0.9	0.7	0.3	0.1	0.243	0.007
0.9	0.1	0.2	0.8	0.1	0.9	0.4	0.6	0.9	0.9	0.1	0.1	0.018	0.072
0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0

Mean A = (0.9 + 0.9 + 0.0) / 3 = 0.6
Product A = 0.9 * 0.9 *0.0 = 0.0

Question 3:

There is no defined percentage in that won't be presented in bootstrap sample, but it is mostly around 37%.
The formula to calculate it is : $1-(1\ -\ 1/N)^N$
For the graph below we used N = [100, 200, 400, 500, 760, 1200,] sample size.



Question 4:

The bootstrapping and random subspace models are pretty same. However, there is a difference between this two, features. While in bootstrapping we create randomly sub-samples of the data set, in RSM we randomly sample the features. So, in bootstrapping we try to avoid the error caused from the small data set and in RSM we try to avoid the focus from the highly predictive features. Beside features, the other difference between bootstrapping and RSM is that bootstrapping is used more when we have small sample as, they are more convenient to mislead the result. On the other hand, RSM is used when we have high number of features but we ignore less predictive features.