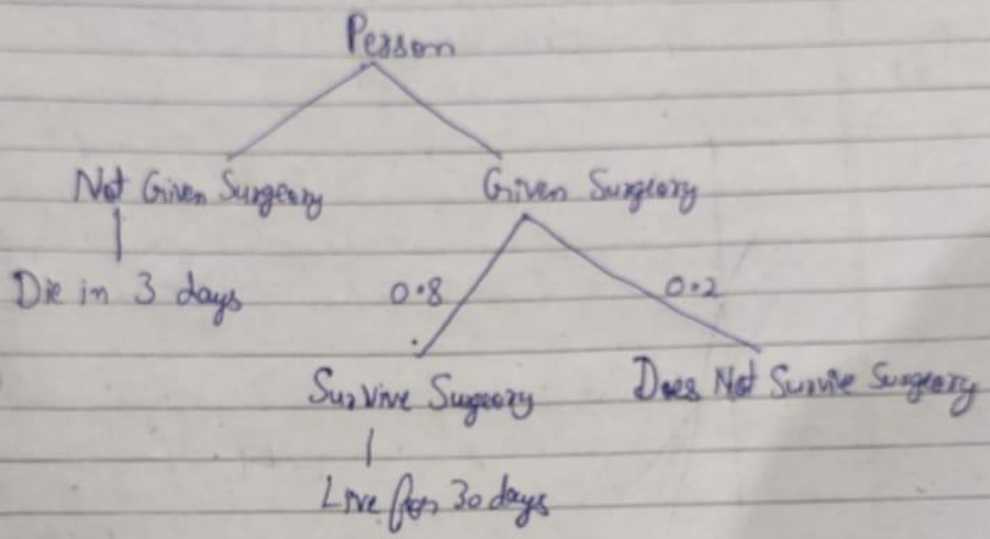
Ans 1:

Ans→

(a)



(b) The patient's utility for living 3 days & have the surgery performed can be quite low, this is because $L(\emptyset) = 0.8$ & it is given that person die if not given surgery in 3 days.

$P(\text{Survive Surgery}) = 0.8$

(c) $P\left(\frac{\text{Positive}}{\text{Survive Surgery}}\right) = 0.95$ $\quad P\left(\frac{\text{Negative}}{\text{Survive Surgery}}\right) = 0.05$

$P\left(\frac{\text{Positive}}{\text{Not Survive Surgery}}\right) = 0.05$ $\quad P\left(\frac{\text{Negative}}{\text{Not Survive Surgery}}\right) = 0.95$

$P(\text{Not Survive Surgery}) = 0.2$

$P\left(\frac{\text{Surviving Surgery}}{\text{Positive}}\right) = ?$

$$P\left(\frac{\text{Surviving Surgery}}{\text{Positive}}\right) = \frac{P\left(\frac{\text{Positive}}{\text{Surviving Surgery}}\right) \times P(\text{Surviving Surgery})}{P\left(\frac{\text{Positive}}{\text{Surviving Surgery}}\right) \times P(\text{Surviving Surgery}) + P\left(\frac{\text{Positive}}{\text{Not Surviving Surgery}}\right) \times P(\text{Not})}$$

$$\Rightarrow \frac{0.95 \times 0.8}{0.95 \times 0.8 + 0.05 \times 0.2} = \frac{0.76}{0.76 + 0.01}$$

$$= \frac{0.76}{0.77}$$

$$= 0.987$$

(d) $P\left(\frac{\text{Not Surviving Surgery}}{\text{Positive}}\right) = 1 - P\left(\frac{\text{Surviving Surgery}}{\text{Positive}}\right)$

$= 1 - 0.987$

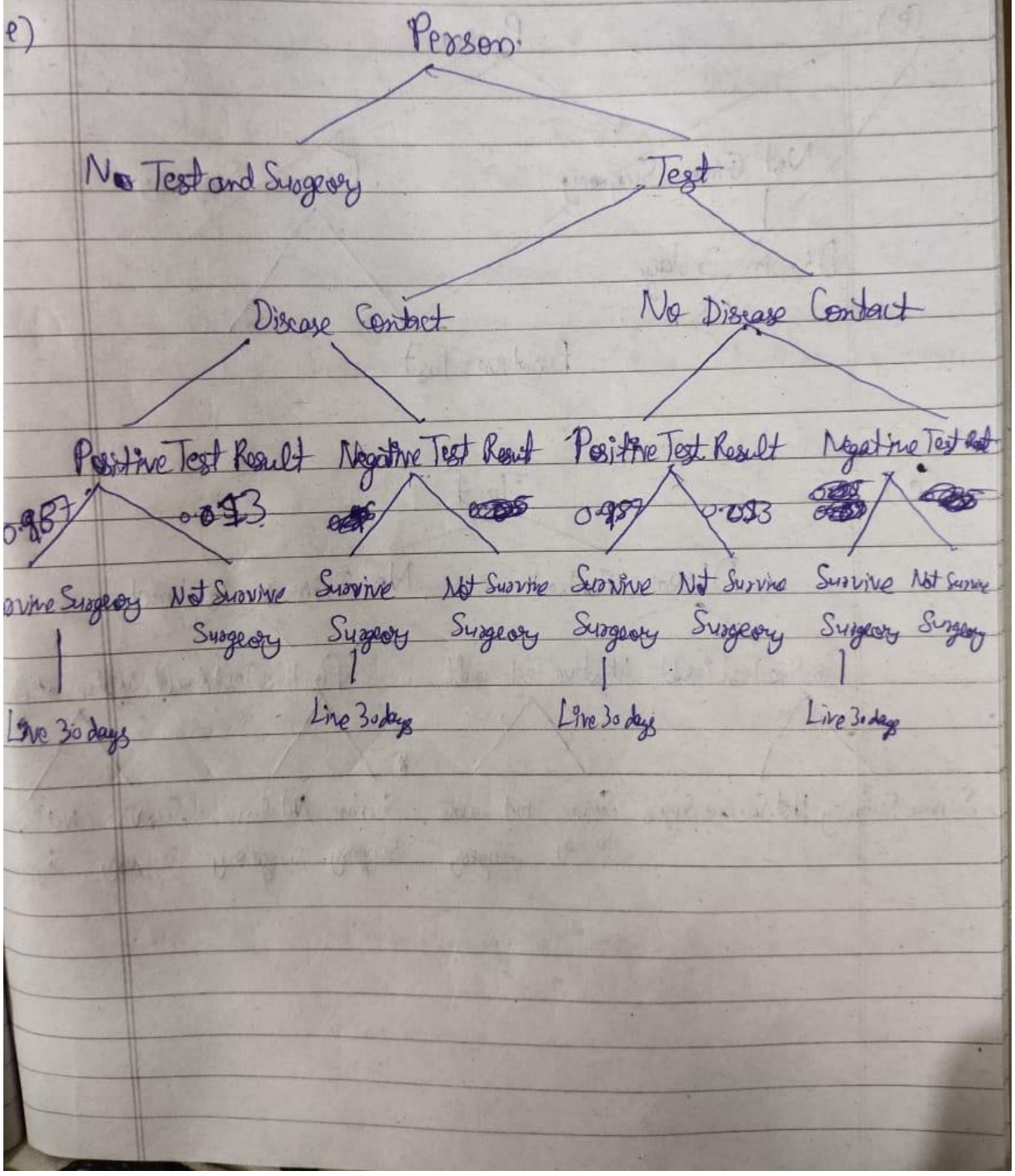
$= 0.013$

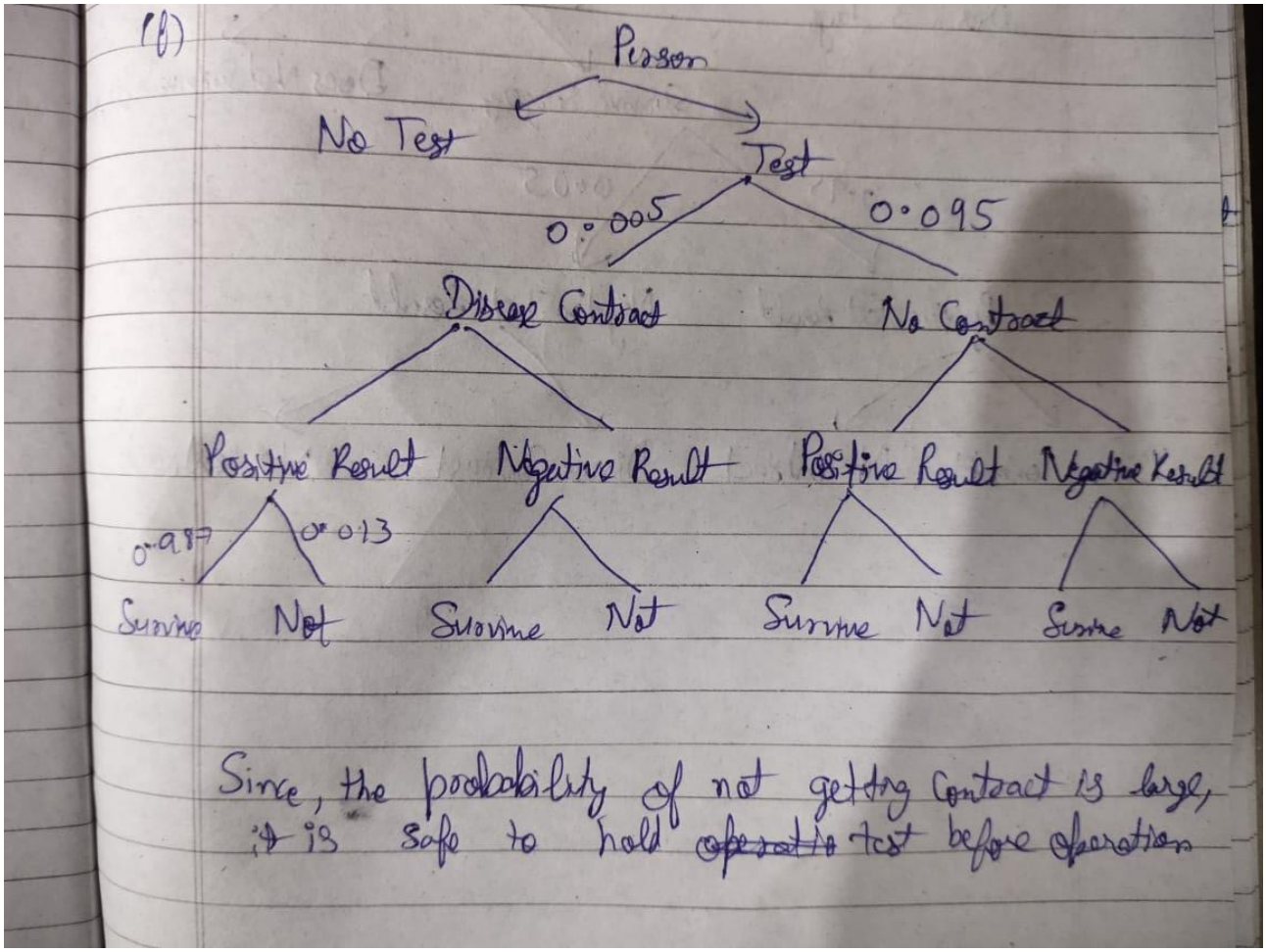
It is clear from above two Probabilities that the Probability of Surviving a Surgery is greater than Not Surviving it when the test is positive.

Thus, Surgery should be performed if test is positive.

Perem

e)

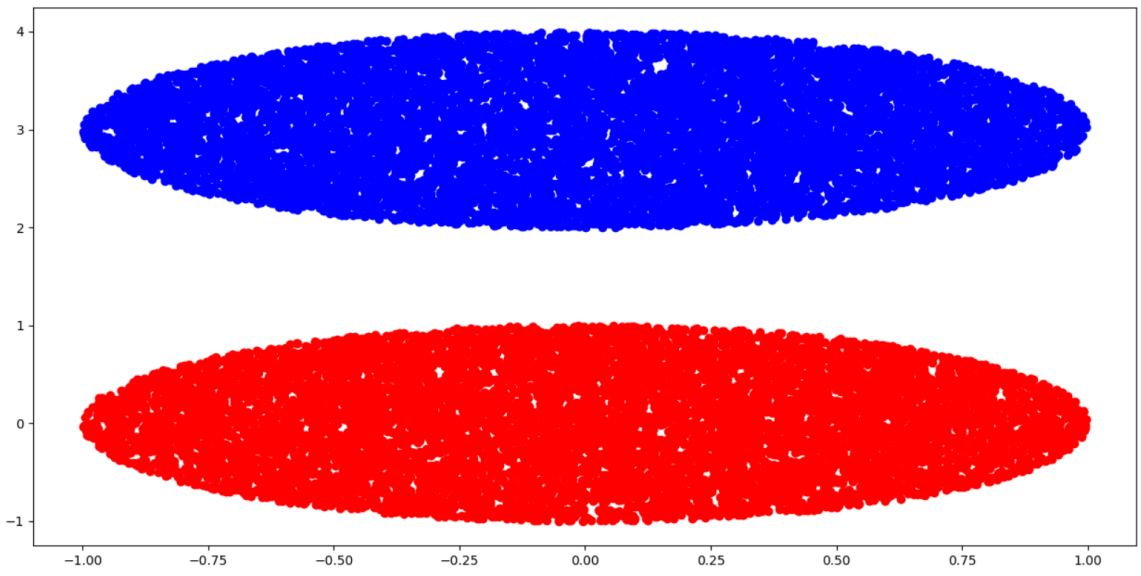

Person
No Test and Surgery
Test
Disease Contact
No Disease Contact
Positive Test Result
Negative Test Result
Positive Test Result
Negative Test Result
Survive Surgery
Not Survive Surgery
Survive Surgery
Not Survive Surgery
Survive Surgery
Not Survive Surgery
Survive Surgery
Not Survive Surgery
Live 30 days
Live 30 days
Live 30 days
Live 30 days



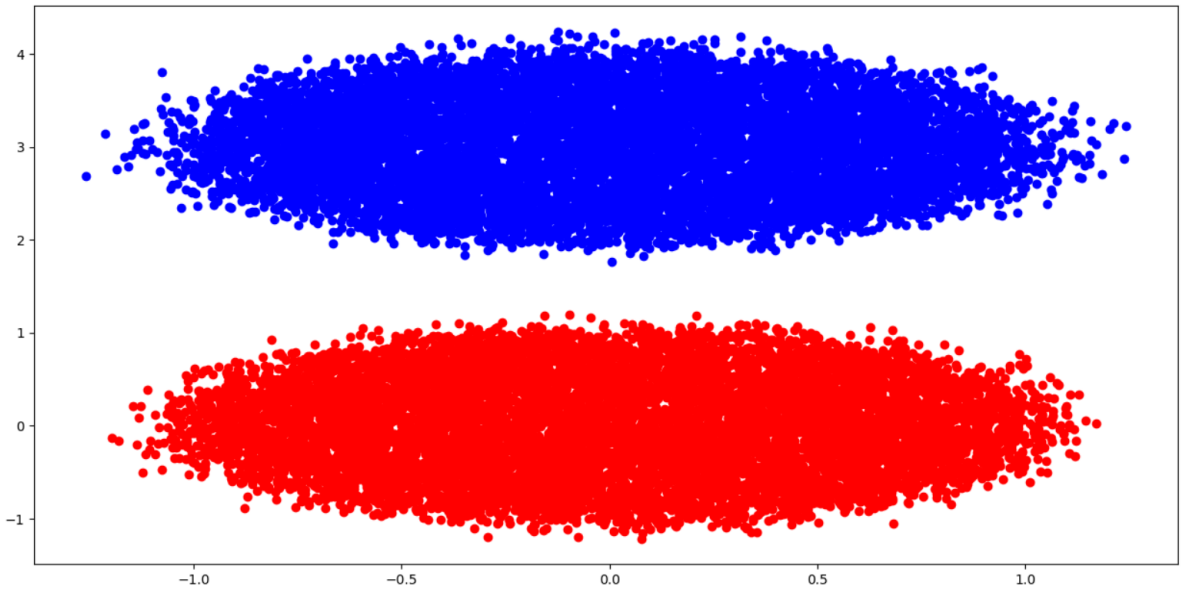
ANS2:

b.

Without Noise

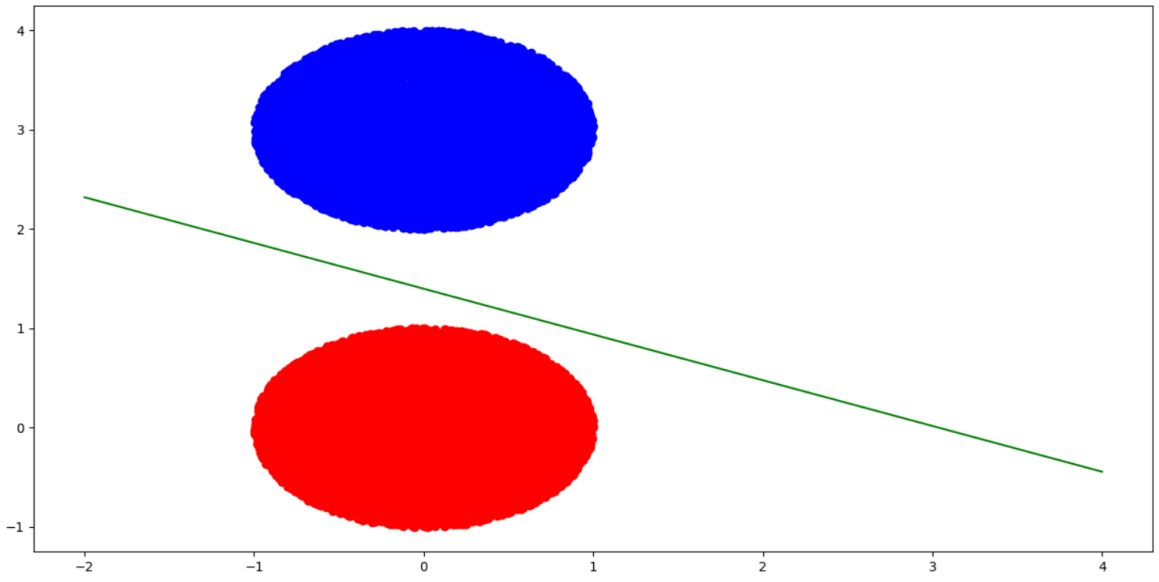


With Noise

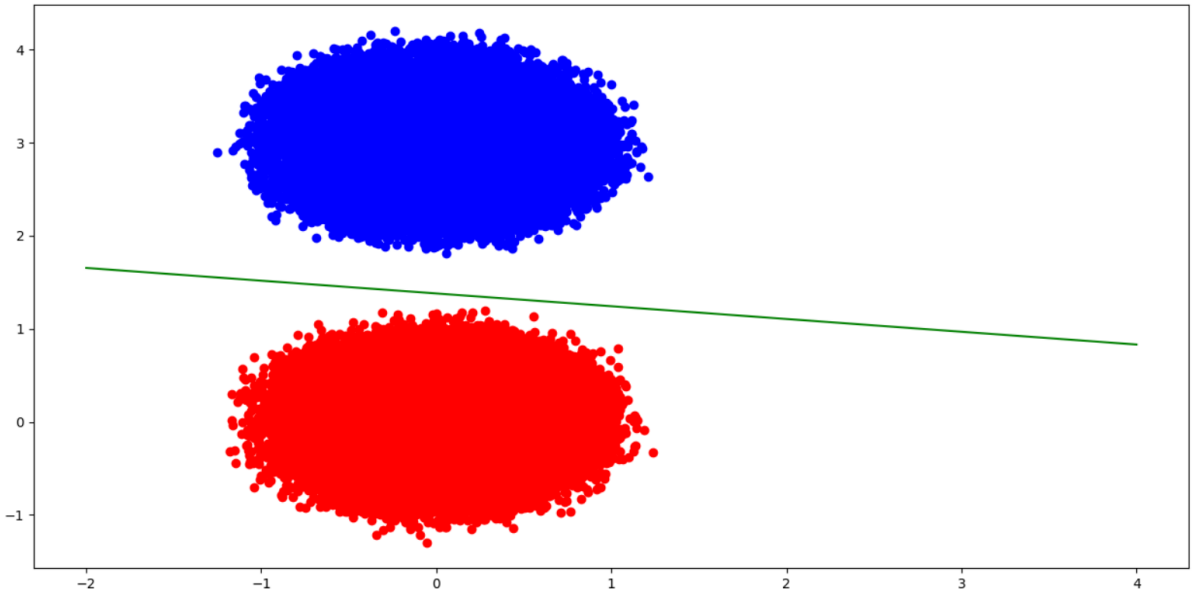


a. In this case there exists a decision boundary which clearly divides both the labels.

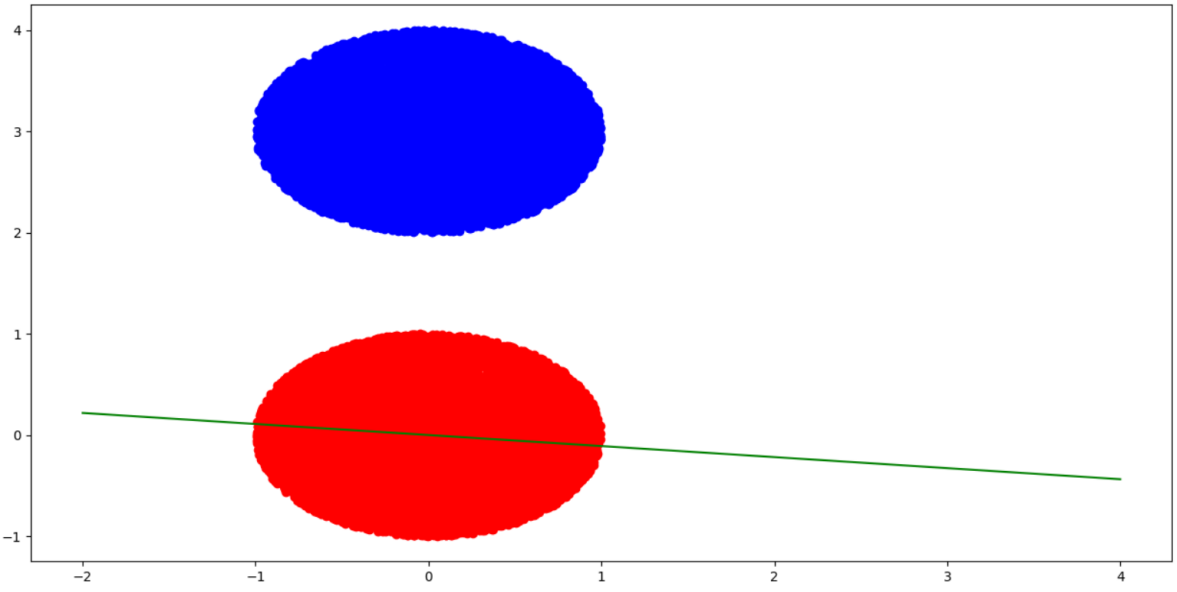
Without Noise:



With Noise:



b. In this case there is no decision boundary that clearly divides the two labels and this is because the bias is here set out to be zero and due to the absence of this bias the model has become constrained about a certain region only. And when the bias is set to be zero the error that the model suffers is also neglected.



c.

AND

X1	X2	Output
0	0	0
0	1	0

1	0	0
1	1	1

OR

X1	X2	Output
0	0	0
0	1	1
1	0	1
1	1	1

XOR

X1	X2	Output
0	0	0
0	1	1
1	0	1
1	1	0

Now when the bias is fixed to be 0 it means the line has to pass the origin only. And because of which there does not exists any decision boundary which could divide the points such that the ones with class 0 are at one side and one with class 1 are on another side of the line.

f.

From a equation of hyperplane which divides the output into two different classes which means one side of the line contain one class and other side of line contain another class. Now to check which side it lies we would put the point in the line and check if the overall value is positive or negative. If it is positive then it lies above the line which means it lies to a particular class but if it is negative it lies below the line which means it belong to the other class.

ANS3:

a.

Criteria	Depth=4	Depth=8	Depth=10	Depth=15	Depth=20
Entropy (Testing Set)	0.9865372967116609	0.9860207311916436	0.9872915738105851	0.9882218488931555	0.9862561570724906
Gini (Testing Set)	0.9857167346658895	0.9863864412978137	0.9868047222317459	0.9880275653992526	0.9867521514003984

Entropy (Validation Set)	0.985536165300968	0.9858790185255025	0.9872390029828231	0.9879932800767991	0.9860893018365504
Gini (Validation Set)	0.985536165300968	0.9862630141369813	0.9867475800276568	0.9879018525502565	0.9866607238774414

From the above table we can see entropy is giving better result thus would be using it further.

```
PS C:\Users\91991> python -u "c:\Users\91991\Desktop\ML\Q3-a.py"
Entropy model score for depth 4 on testing data is : 0.9857167346658895
Gini model score for depth 4 on testing data is : 0.9857167346658895
Entropy model score for depth 4 on validation data is : 0.985536165300968
Gini model score for depth 4 on validation data is : 0.985536165300968
Entropy model score for depth 8 on testing data is : 0.9860207311916436
Gini model score for depth 8 on testing data is : 0.9863864412978137
Entropy model score for depth 8 on validation data is : 0.9858790185255025
Gini model score for depth 8 on validation data is : 0.9862630141369813
Entropy model score for depth 10 on testing data is : 0.9872915738105851
Gini model score for depth 10 on testing data is : 0.9868047222317459
Entropy model score for depth 10 on validation data is : 0.9872390029828231
Gini model score for depth 10 on validation data is : 0.9867475800276568
Entropy model score for depth 15 on testing data is : 0.9882218488931555
Gini model score for depth 15 on testing data is : 0.9880275653992526
Entropy model score for depth 15 on validation data is : 0.9879932800767991
Gini model score for depth 15 on validation data is : 0.9879018525502565
Entropy model score for depth 20 on testing data is : 0.9862561570724906
Gini model score for depth 20 on testing data is : 0.986752151403984
Entropy model score for depth 20 on validation data is : 0.9860893018365504
Gini model score for depth 20 on validation data is : 0.9866607238774414
```

b. Here the accuracy is coming out to be 0.9797162836419099. The value is less than that obtained from the above part and this is because the max depth is set out to be 3 and because of which stumps are acting like weak classifiers.

c.

Technique	Estimator=4	Estimator=8	Estimator=10	Estimator=15	Estimator=20
Adaboost	0.9846813179277951	0.9854744517205518	0.9845076056273643	0.9843270362624427	0.9851247414315265
Random Forest	0.9850401709694746	0.9873052879395664	0.9878835670449481	0.9882309916458097	0.9881029931086502

Overall we can see that random forest is a better performer than adaboost.