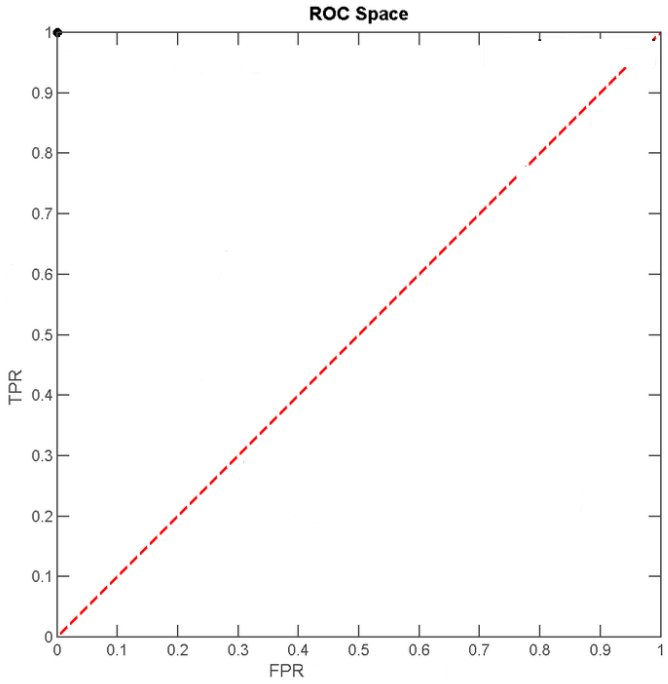
Name: Islam,Rafiqul Id:17-34438-1 Sec:A

Classifiers

Four classifiers are generated for the same training set, which has 100 instances. They have the following confusion matrices.

Classifier 1:

	1	2	3	4	5	6
1	10	4	4	2	0	0
2	1	10	2	0	3	0
3	10	0	10	0	1	0
4	0	1	1	15	10	1
5	0	0	4	0	20	1
6	0	0	0	0	5	1



Classifier 2:

	1	2	3	4	5	6
1	0	4	4	2	0	0
2	1	8	2	1	3	0
3	0	2	10	2	1	0
4	0	1	1	1	10	1
5	0	0	0	4	12	1
6	0	0	0	1	0	10

Which classifier would you consider the best if you want to develop a cancer patient identification classifier? Note that your +ve interest is to admit cancer stage 1, 3 and 5. Use ROC Graph for visualization.

Answer:

So, we got 1,3,5 is positive for cancer stage

And 2,4,6 is negative for cancer stage

So, if we swap column and row in Classifier 1 and Classifier 2 for our calculation advantage, we get

Classifier 1:

	1	3	5	2	4	6
1	10	4	0	4	2	0
3	10	10	1	0	0	0
5	0	4	20	0	0	1
2	1	2	3	10	0	0
4	0	1	10	1	15	1
6	0	0	5	0	0	1

If we put the data into confusion matrix then,

		Predicted Class		
		+	-	**Total**
Actual Class	**+**	59	7	66
	-	22	28	50
	Total	81	35	116

So, we get,

TPR = $\frac{59}{66}$ = 0.80 and FPR = $\frac{22}{50}$ = 0.44

Classifier 2:

	1	3	5	2	4	6

1	0	4	0	4	2	0
3	0	10	1	2	2	0
5	0	0	12	0	4	1
2	1	2	3	8	1	0
4	0	1	10	1	1	1
6	0	0	0	0	1	10

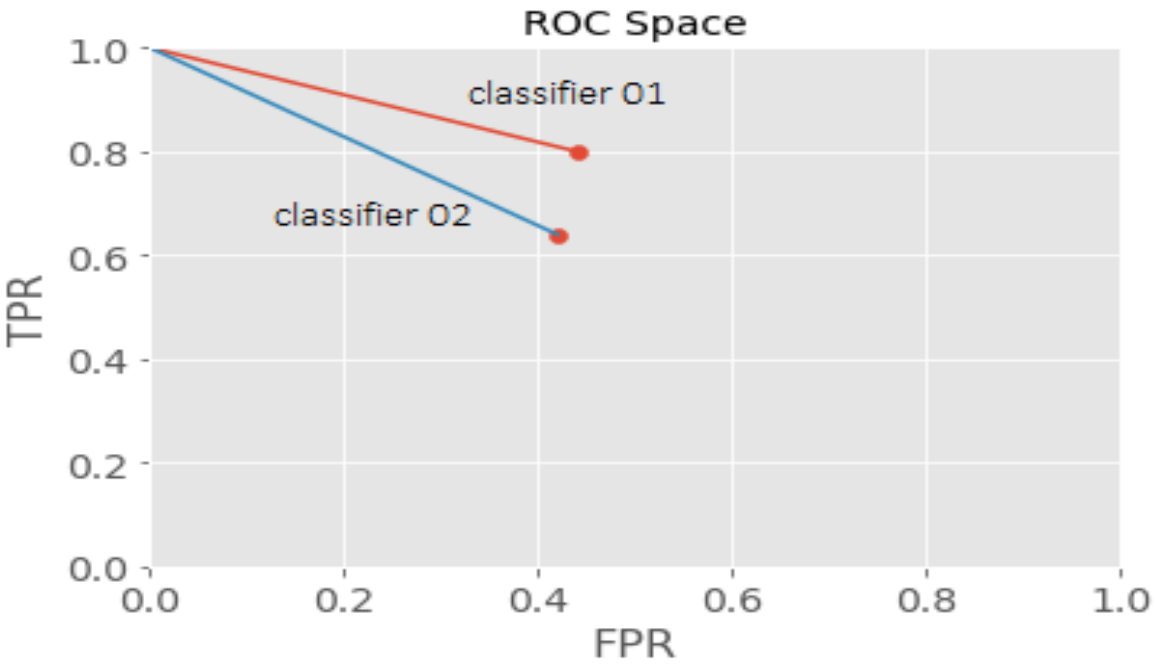
If we put the data into confusion matrix then,

		Predicted Class		
		+	-	**Total**
Actual Class	**+**	27	15	42
	-	17	23	40
	Total	44	38	82

So, we get,

TPR = $\frac{27}{42}$ = 0.64 and FPR = $\frac{17}{40}$ = 0.42

And, if we plot this points on graphs we get,



And if we consider best point (0,1) and calculate distance from the best point to classifier 01 point and classifier 02 point

Distance between Best_point and Classifier01_point:

Distance 01 = $\sqrt{(0-.44)^2+(1-.80)^2}$ = 0.483

Distance between Best_point and Classifier02_point:

Distance 02 = $\sqrt{(0-.42)^2+(1-.64)^2}$ = 0.553

We can see Distance 01 is less than Distance 02 ,we can say that Classifier 01 will work better than Classifier 02