



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

# Assignment 1 (Final)

## Information Retrieval Applications

A web search engine can be treated as a kind of classifier.

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

		Predicted class	
		+	-
A	Actual class = +	25	15
	Actual class = -	25	35
B	Actual class = +	35	15
	Actual class = -	15	35
C	Actual class = +	25	25
	Actual class = -	15	35
D	Actual class = +	35	25
	Actual class = -	25	15

Which classifier would you consider the best if you were equally concerned with avoiding false positive and false negative classifications?

For point A

$$FPR = \frac{10}{40} = .25$$

$$TPR = \frac{50}{60} = .83$$

$$FPR = \frac{5}{40} = .125$$

$$TPR = \frac{35}{60} = .58$$

③ distance from best case

$$\sqrt{(0 - .25)^2 + (1 - .83)^2} = .229$$

$$④ \sqrt{(0 - .125)^2 + (1 - .58)^2} = .16$$

$$⑤ \sqrt{(0 - .025)^2 + (1 - .5)^2} = .501$$

$$⑥ \sqrt{(0 - .2)^2 + (1 - .0)^2} = .2$$

$$① FPR = \frac{1}{40} = .025$$

$$TPR = \frac{40}{60} = .67$$

$$② FPR = \frac{20}{40} = .5$$

$$TPR = \frac{60}{60} = 1$$

ROC curve is a graphical representation of the performance of a binary classifier system. The curve is created by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR).



• Finding the value inside ROC curve is one step ahead. It is the measure of how well the classifier is performing. The area under the curve is the measure of the classifier's performance. The area under the curve is the measure of the classifier's performance. The area under the curve is the measure of the classifier's performance.

