Computational Vision

Lecture 5.2: ROC Analysis

Hamid Dehghani

Office: UG38

Receiver operating characteristic

- ROC analysis provides tools to select possibly optimal models and to discard suboptimal ones
- first developed by electrical engineers and radar engineers during the World War II for detecting enemy objects in battle fields
- widely used in medicine, radiology, psychology and other areas for many decades, it has been introduced relatively recently in other areas like machine learning and data mining.

Evaluating vision algorithms

- You have designed a new edge detection technique.
- You give it to me, and I try it on my image dataset where the task is to predict whether the scene contains a chair or not.
- I tell you that it achieved 95% accuracy on my data.
- Is your technique a success?

Types of errors

- But suppose that
  - The 95% is the correctly classified pixels
  - Only 5% of the pixels are actually edges
  - It misses all the edge pixels

- How do we count the effect of different types of error?

Types of errors

Ground Truth		Prediction: **Edge**	Prediction: **Not edge**
	Edge	True Positive	False Negative
	Not Edge	False Positive	True Negative

Two parts to each: whether you got it correct or not, and what you guessed. For example for a particular pixel, our guess might be labelled…

True Positive

Did we get it correct?
True, we did get it correct.

What did we say?
We said ‘positive’, i.e. edge.

or maybe it was labelled as one of the others, maybe…

False Negative

Did we get it correct?
False, we did not get it correct.

What did we say?
We said ‘negative, i.e. not edge.

Sensitivity and Specificity

Count up the total number of each label (TP, FP, TN, FN) over a large dataset. In ROC analysis, we use two statistics:

Sensitivity = $\frac{TP}{TP+FN}$

Can be thought of as the likelihood of spotting a positive case when presented with one.

Or… the proportion of edges we find.

Specificity = $\frac{TN}{TN+FP}$

Can be thought of as the likelihood of spotting a negative case when presented with one.

Or… the proportion of non-edges that we find

$$\text{Sensitivity} = \frac{TP}{TP+FN} = ?$$

$$\text{Specificity} = \frac{TN}{TN+FP} = ?$$

		Prediction	
		1	**0**
Ground Truth	**1**	60	30
	0	80	20

60+30 = 90 cases in the dataset were class 1 (edge)

80+20 = 100 cases in the dataset were class 0 (non-edge)

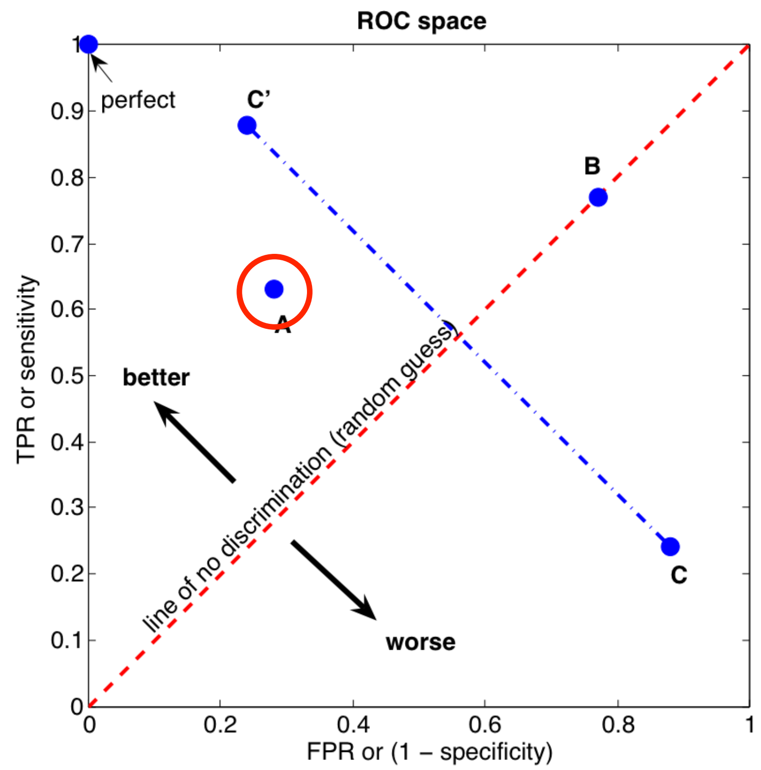
90+100 = 190 examples (pixels) in the data overall

The ROC space

- only the true positive rate (TPR) and false positive rate (FPR) are needed.
  - TPR determines a classifier test performance on classifying positive instances correctly among all positive samples available during the test (sensitivity).
  - FPR defines how many incorrect positive results occur among all negative samples available during the test (1-specificity).
- A ROC space is defined by FPR and TPR as x and y axes respectively

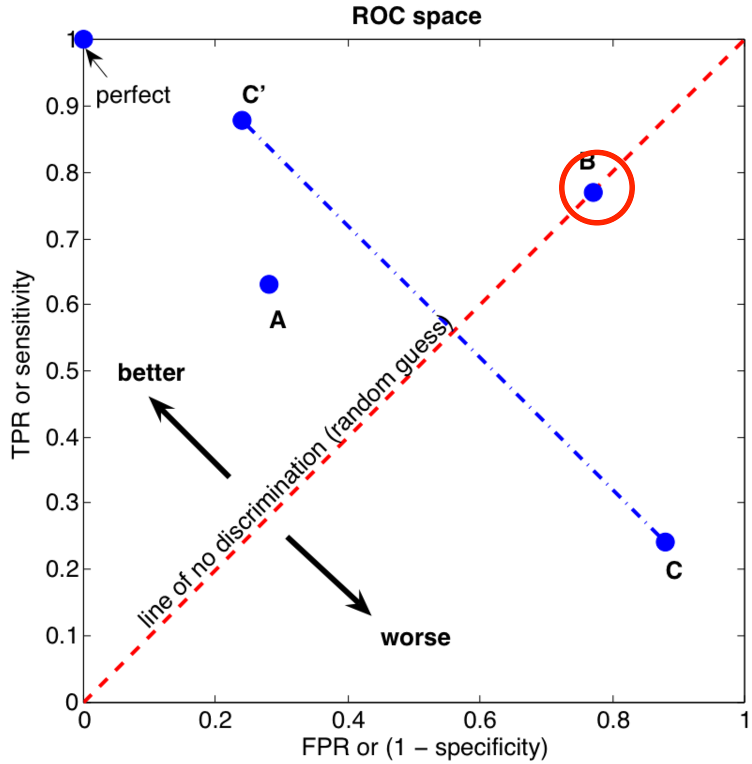
The ROC space

		Prediction 1	Prediction 0
Ground Truth	1	63	37
	0	28	72



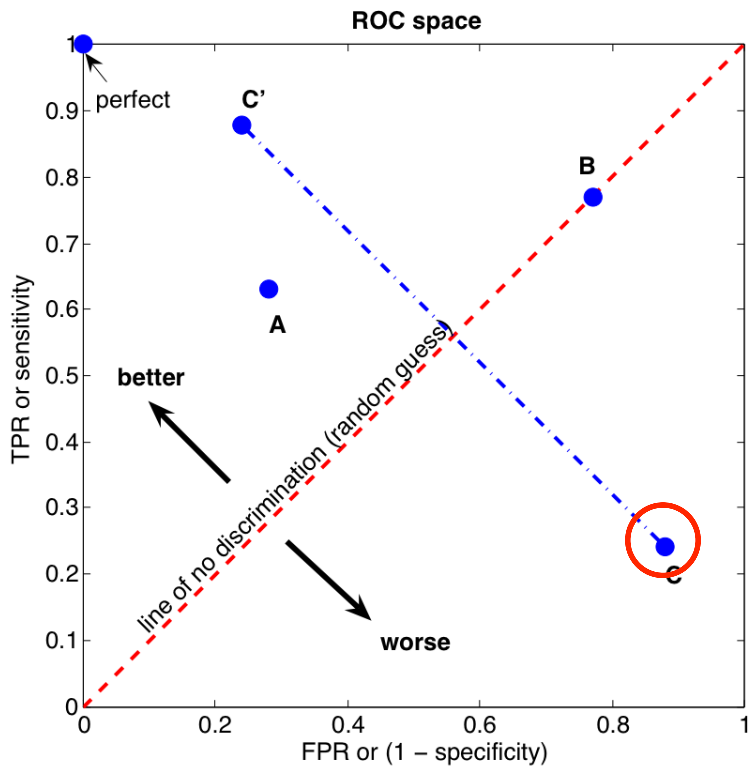
The ROC space

		Prediction	
		1	0
Ground Truth	1	77	23
	0	77	23

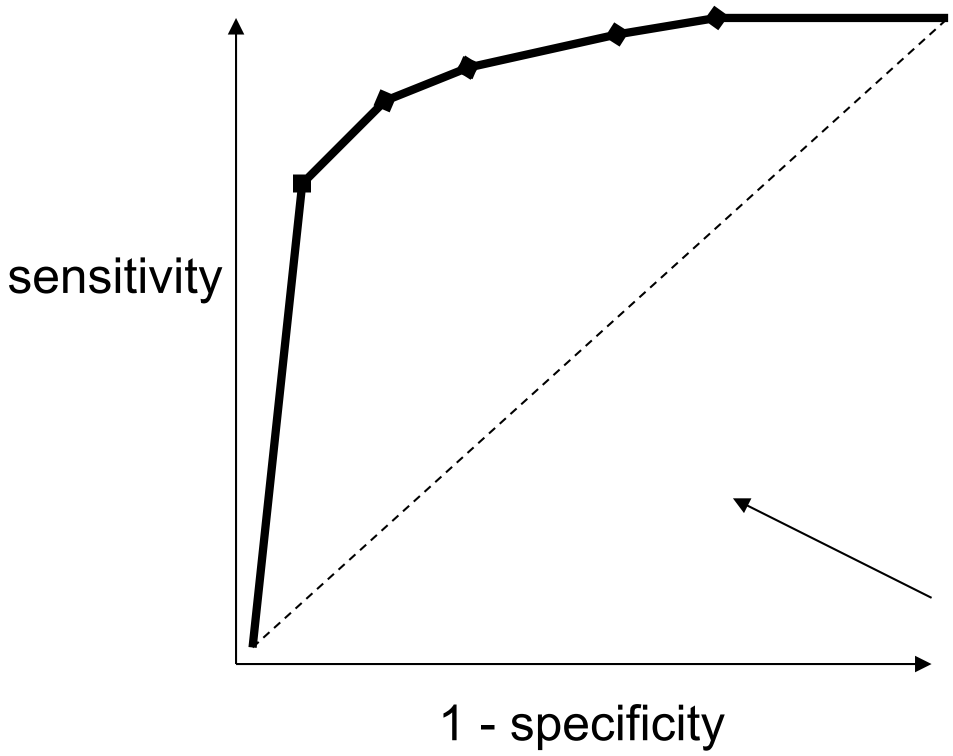


The ROC space

		Prediction	
		1	**0**
Ground Truth	**1**	24	76
	0	88	12



ROC Analysis



All the optimal detectors lie on the convex hull.

Which of these is best depends on the ratio of edges to non-edges, and the different cost of misclassification

Any detector on this side can lead to a better detector by flipping its output.

<u>Take-home point :</u> You should always quote sensitivity and specificity for your algorithm, if possible plotting an ROC graph. Remember also though, <u>any</u> statistic you quote should be an average over a suitable range of tests for your algorithm.