CSE 435 Pattern Recognition

ROC curves

Tanvir Azhar
Lecturer, EDU

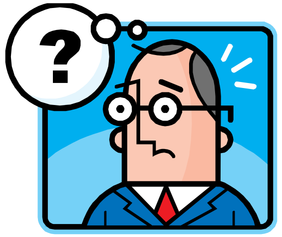
No, it is not “Rock’n’Roll!”

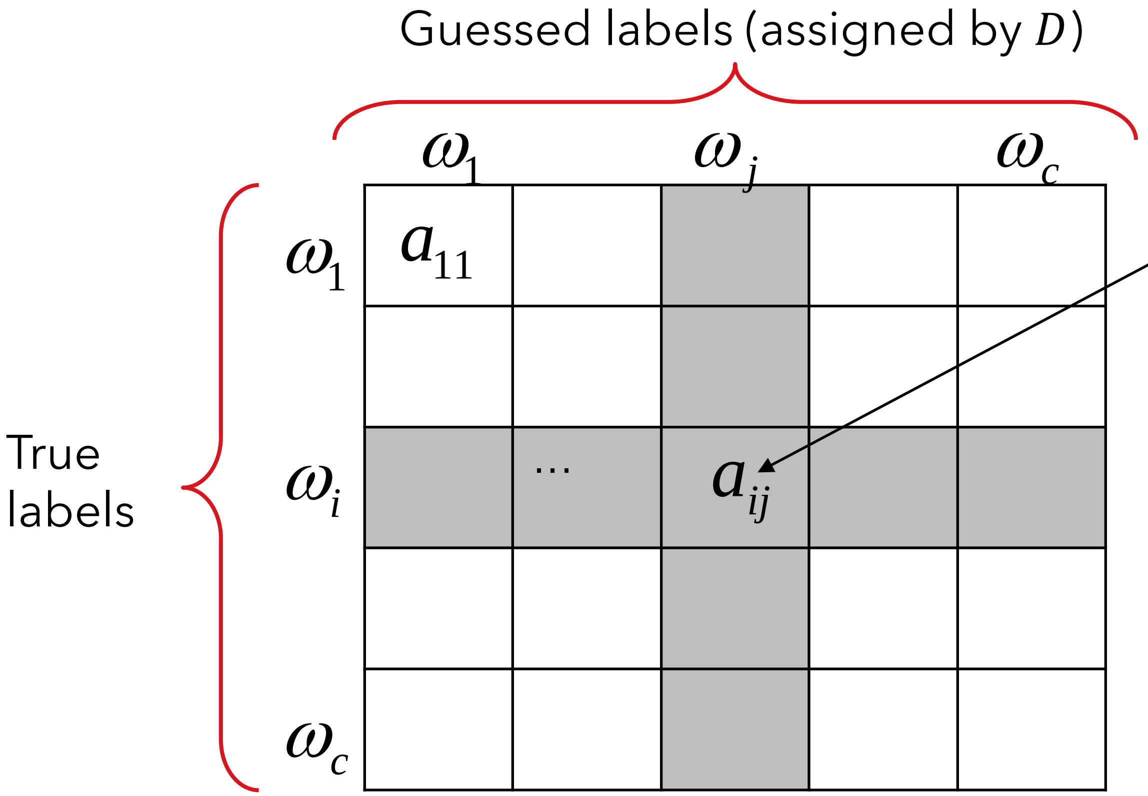
ROC curves

(Receiver Operating Characteristics)

- Positive and negative classes. Type I and Type II errors
- Sensitivity, Specificity, Recall, Precision
- Why is accuracy not sufficient? F1 measure
- Face detection example
- ROC curves

Recall: Confusion matrix





The number of elements of the testing set whose true label is ω_i and D has assigned them to ω_j

Positives and negatives for a binary TEST

Assigned

True

	Guilty (+)	Innocent (-)
Guilty (+)	TP Convict the guilty	FN Free the guilty
Innocent (-)	FP Convict the innocent	TN Free the innocent

Positives and negatives for a binary TEST

Assigned

True

	Disease (positive)	Healthy (negative)
Disease (positive)	TP (true positives)	FN (false negatives)
Healthy (negative)	FP (false positives)	TN (true negative)

Positives and negatives for a binary TEST

Assigned

True

	Disease (positive)	Healthy (negative)
Disease (positive)	TP (true positives)	**TYPE I ERROR**
Healthy (negative)	**TYPE II ERROR**	TN (true negative)

Positives and negatives for a binary TEST

True	Assigned: Disease (positive)	Assigned: Healthy (negative)
Disease (positive)	TP (true positives)	FN (false negatives)
Healthy (negative)	FP (false positives)	TN (true negative)

Sensitivity of the test = TP / (TP + FN)

Specificity of the test = TN / (TN + FP)

Recall = Sensitivity of the test = TP / (TP + FN)

Precision = TP / (TP + FP)

Accuracy of the test = (TP + TN) / (TP + TN + FP + FN)

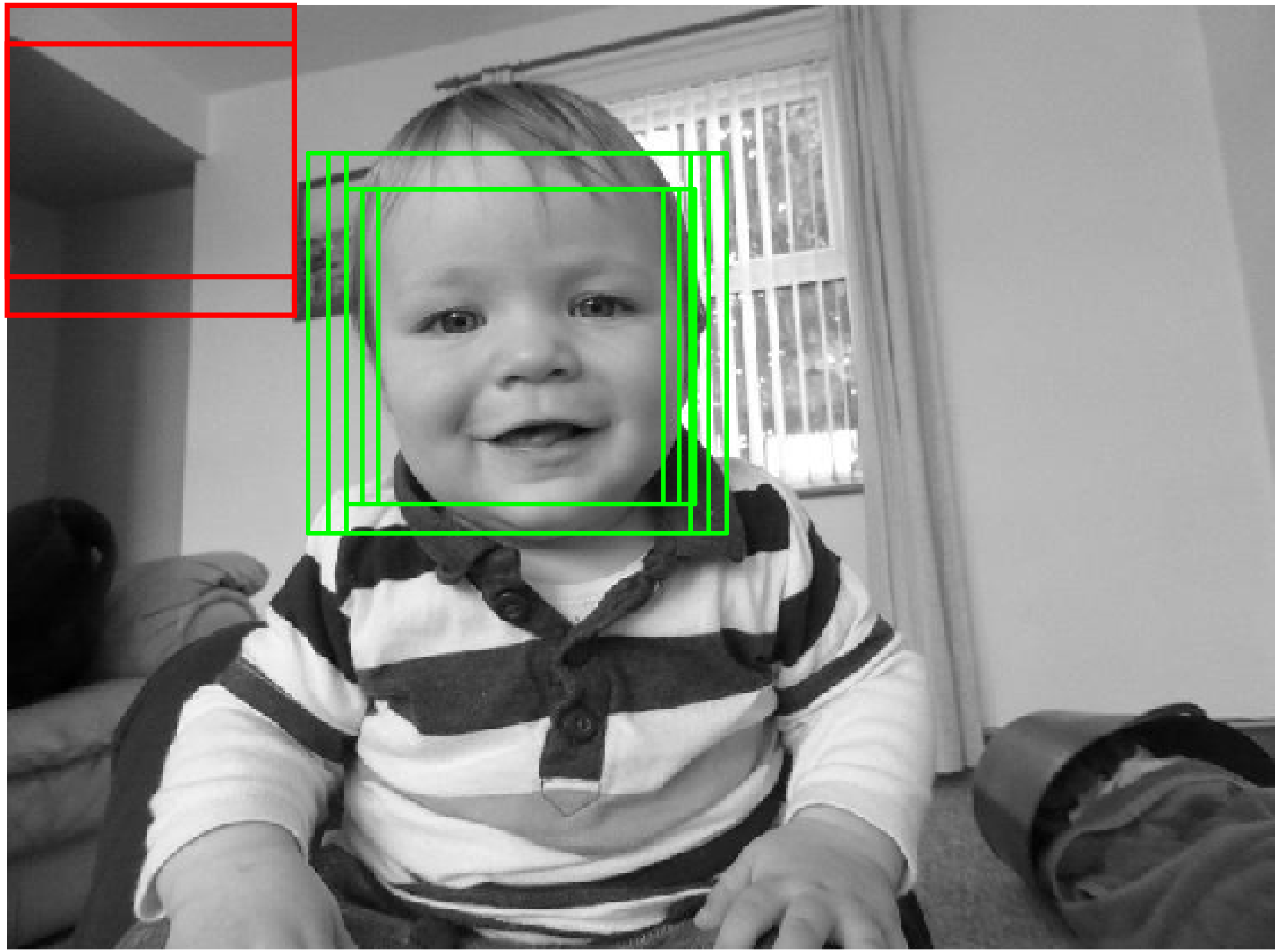
Why is ACCURACY not sufficient?

Sometimes the number of negative cases is much greater than the number of positive cases. This is the usual case in medical screening for a rare disease. **UNBALANCED CLASS PROBLEM**

Suppose there are 1000 cases, 995 of which are negative cases and 5 of which are positive cases. If the system classifies them all as negative, the accuracy would be 99.5%, even though the classifier missed all positive cases. An alternative performance measure that accounts for this is based on the harmonic mean of precision and recall

$$\boxed{F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}}$$

Example of unbalanced classes: face detection



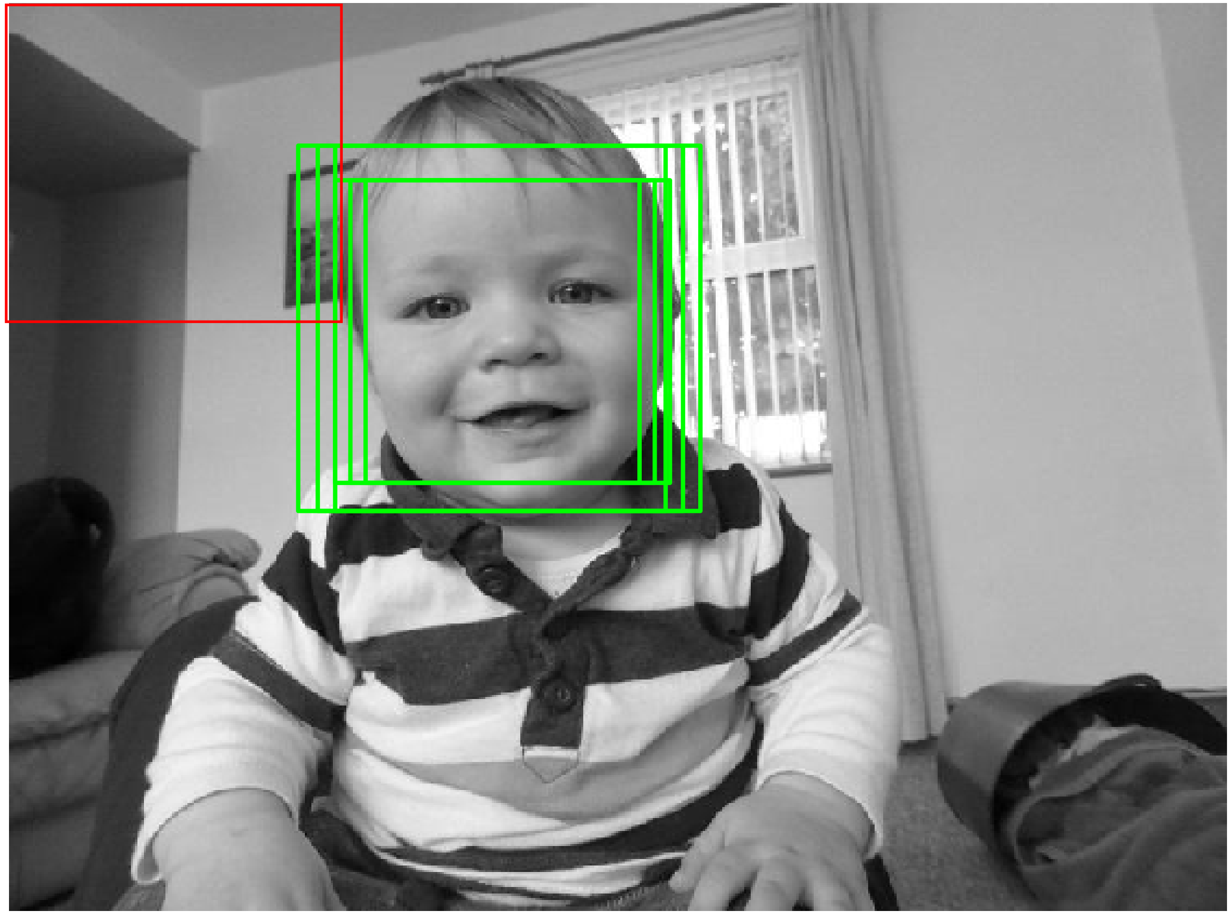
Prepare a data set from this image

c = 2 classes:
“face” and “no face”

objects = squares of a fixed size cropped from the image

features = any set of features extracted form the square

Example of unbalanced classes: face detection



Square 25x25

Image size:
79x118

Positives: 6

Q1.
How many negatives are there?

Suppose that we have a data set (obtained from images with faces) and a classifier D_{Proper}.

Assigned

True

	Face (positive)	No face (negative)
Face (positive)	96	9
No face (negative)	3,505	99,790

Total number of objects, $N = 96 + 9 + 3{,}505 + 99{,}790 = 103{,}400$

Sensitivity, $Sens = \frac{96}{105} = 0.9143$

Specificity, $Spec = \frac{99{,}790}{103{,}295} = 0.9659$

Accuracy

$$Acc = \frac{96 + 99{,}790}{103{,}400} = \mathbf{0.9660}$$

Suppose that we have a data set (obtained from images with faces) and a classifier D_{Proper}.

Assigned

True	Face (positive)	No face (negative)
Face (positive)	96	9
No face (negative)	3,505	99,790

Total number of objects, $N = 96 + 9 + 3{,}505 + 99{,}790 = 103{,}400$

$\text{Recall} = \frac{96}{105} = 0.9143$

$\text{Precision} = \frac{96}{3{,}601} = 0.0267$

F1 measure

$$F_1 = 2 \times \frac{0.9143 \times 0.0267}{0.9143 + 0.0267} = \mathbf{0.0519}$$

Now suppose that we have a classifier $D_{Negative}$ which **always** predicts "no face"

Assigned

True	Face (positive)	No face (negative)
Face (positive)	0	105
No face (negative)	0	103,295

Total number of objects, $N = 96 + 9 + 3{,}505 + 99{,}790 = 103{,}400$

Sensitivity, $Sens = \frac{0}{105} = 0$

Specificity, $Spec = \frac{103{,}295}{103{,}295} = 1$

Accuracy

$$Acc = \frac{103{,}295}{103{,}400} = \mathbf{0.9990}$$

Now suppose that we have a classifier $D_{Negative}$ which **always** predicts "no face"

Assigned

True	Face	No face
Face (positive)		
No face (negative)	0	103,295

Is this a better classifier then?

Total number of objects, $N = 96 + 9 + 3{,}505 + 99{,}790 = 103{,}400$

Sensitivity, $Sens = \frac{0}{105} = 0$

Specificity, $Spec = \frac{103{,}295}{103{,}295} = 1$

Accuracy

$$Acc = \frac{103{,}295}{103{,}400} = \mathbf{0.9990}$$

Now suppose that we have a classifier $D_{Negative}$ which **always** predicts "no face"

Assigned

True	Face (positive)	No face (negative)
Face (positive)	0	105
No face (negative)	0	103,295

Total number of objects, $N = 96 + 9 + 3{,}505 + 99{,}790 = 103{,}400$

$\text{Recall} = \frac{0}{105} = 0$

$\text{Precision} =$ does not exist because there are no positive labels. But even if there were, and even if precision was equal to 1, $F_1 = 2 \times \frac{0 \times 1}{0+1} = 0$.

Compare:

D_{Proper}

Accuracy

$Acc = \mathbf{0.9660}$

$\boldsymbol{F_1 = 0.0519}$

$D_{Negative}$

Accuracy

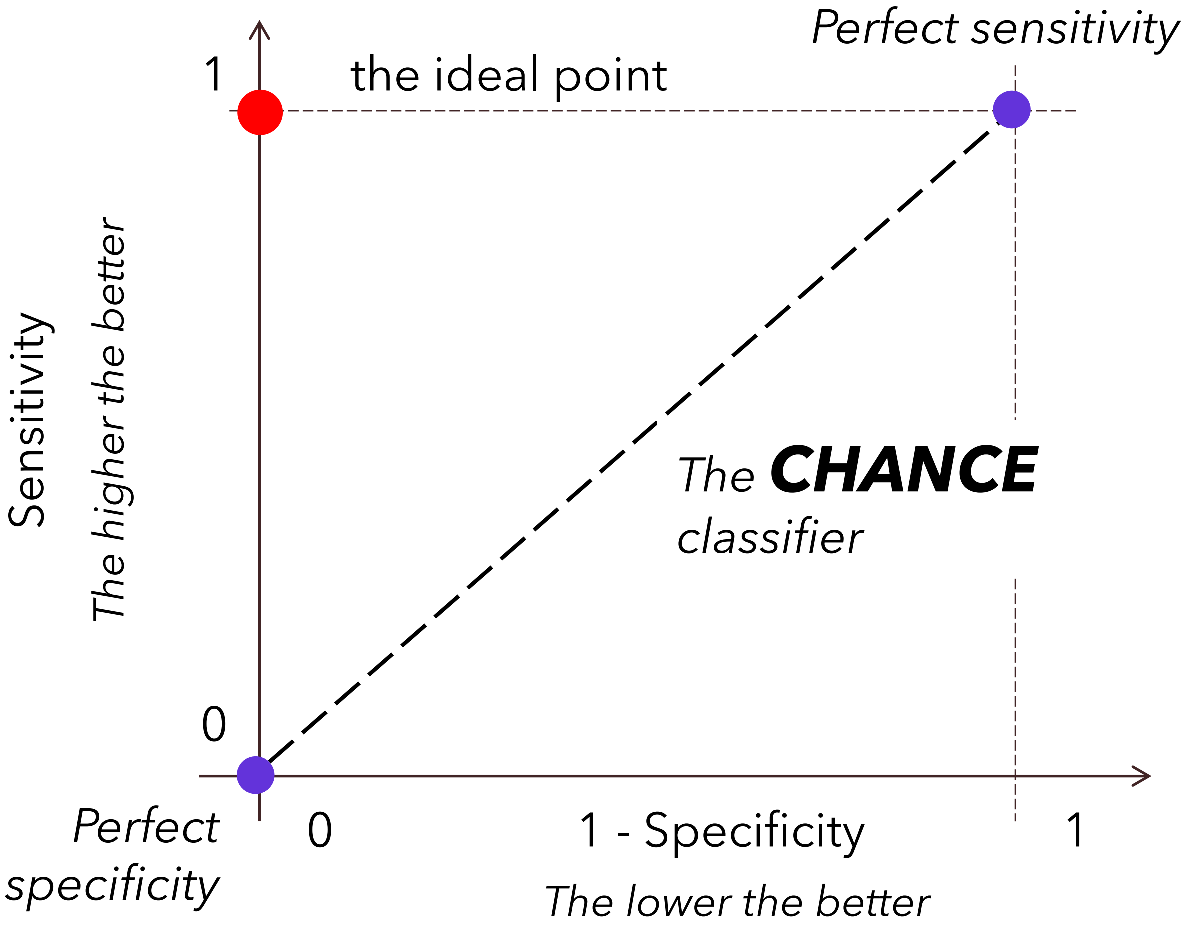
$Acc = \mathbf{0.9990}$

$\boldsymbol{F_1 = 0.0000}$

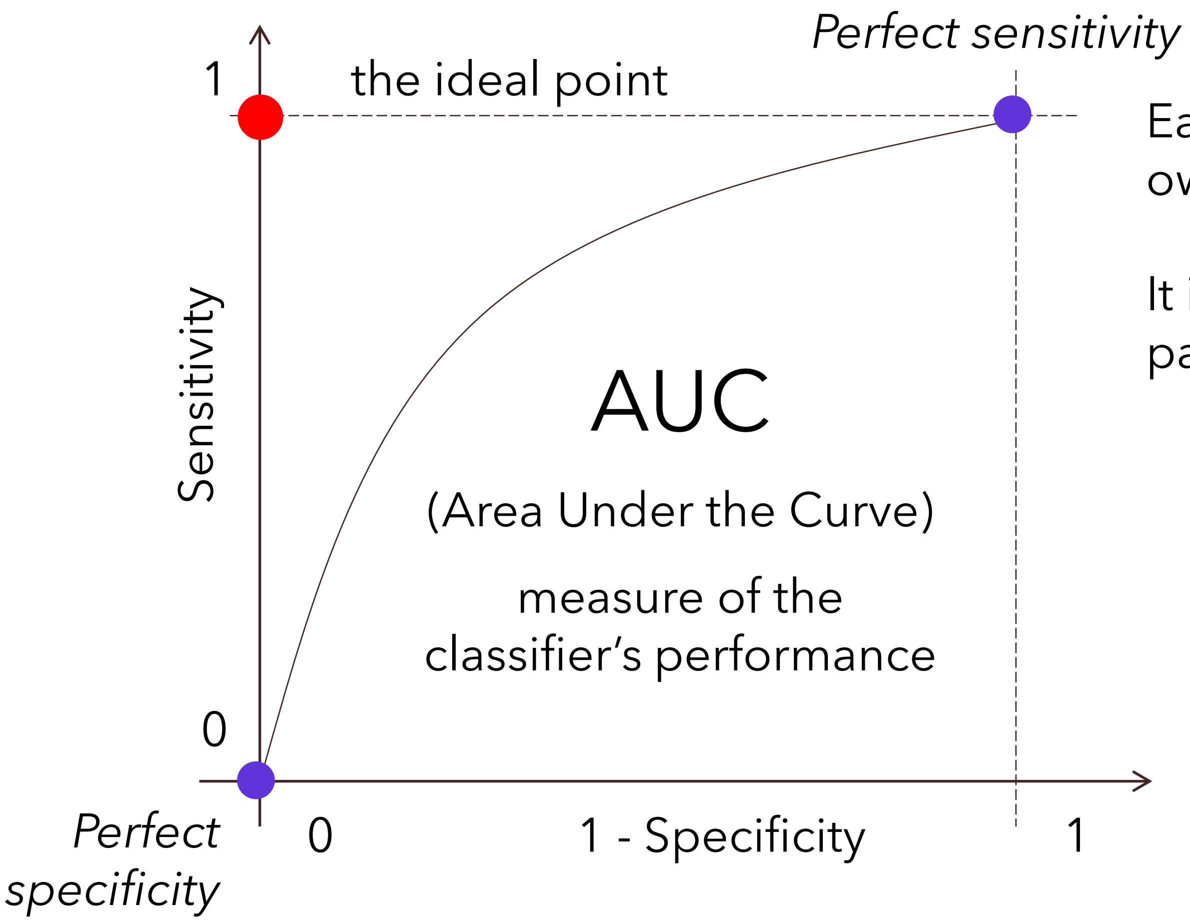
ROC curves
(**R**eceiver **O**perating **C**haracteristic)

ROC curve is a graphical plot that illustrates the performance of a binary classifier (two classes).

ROC curves
(**R**eceiver **O**perating **C**haracteristic)



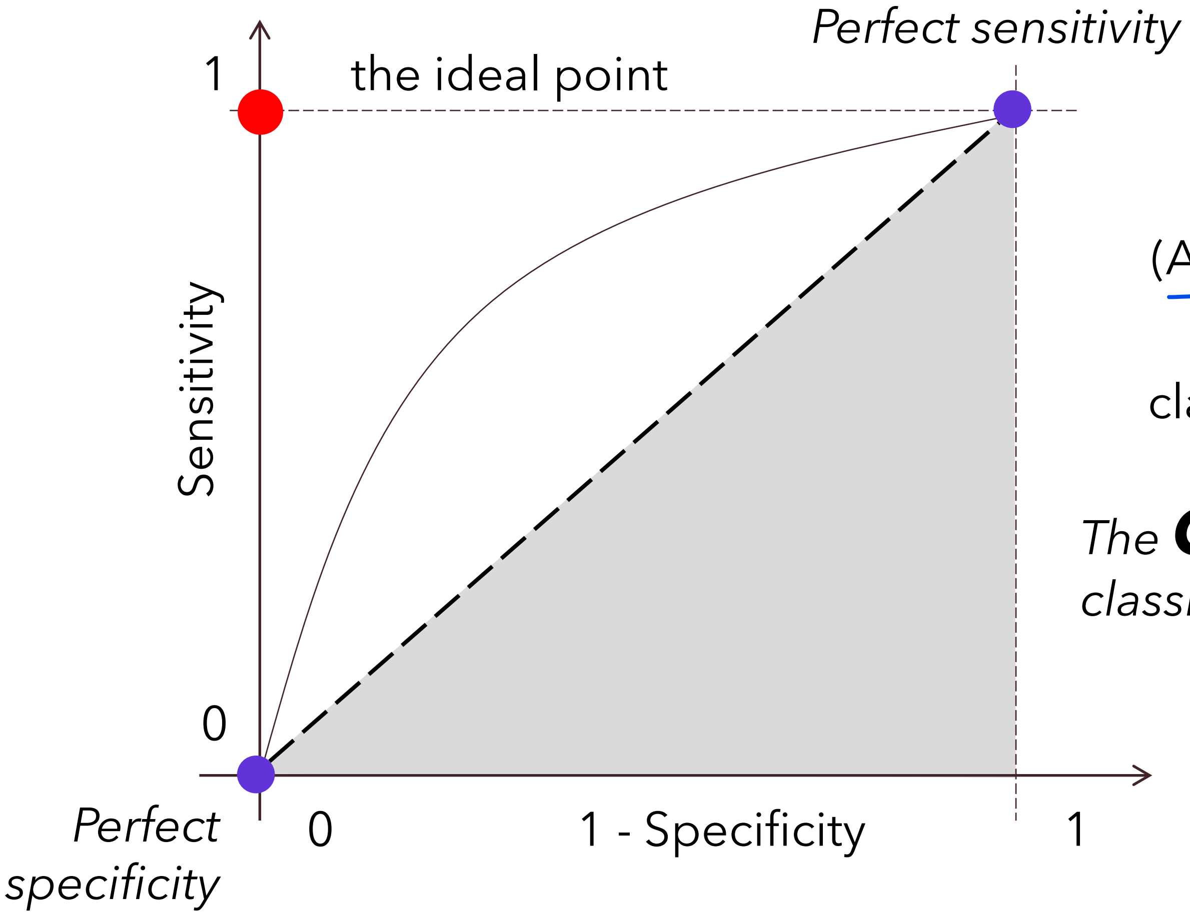
ROC curves
(**R**eceiver **O**perating **C**haracteristic)



Each classifier will have its own ROC curve.

It is obtained by varying a parameter.

ROC curves
(**R**eceiver **O**perating **C**haracteristic)



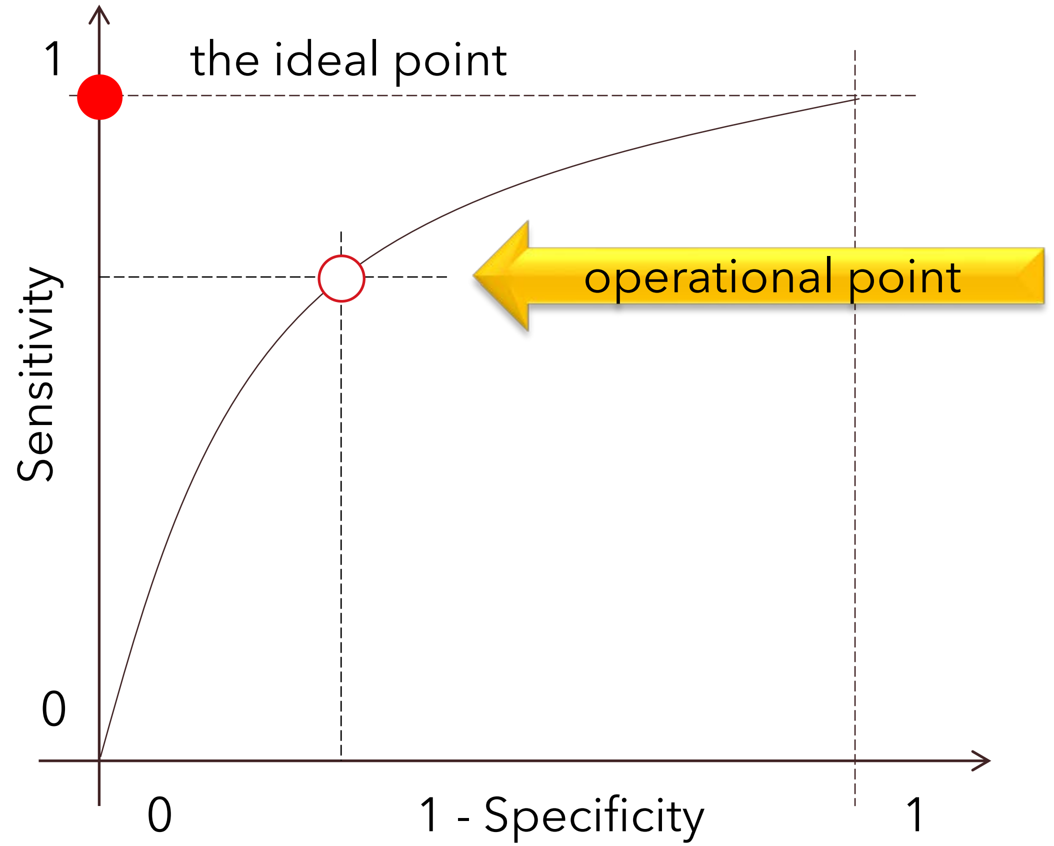
AUC

(Area Under the Curve)

measure of the
classifier's performance

*The **CHANCE***
classifier has AUC = 0.5

ROC curves
(**R**eceiver **O**perating **C**haracteristic)



	Disease +	Healthy -
Disease +	10	4
Healthy -	8	13

Sensitivity = 10/14 = **71.4%**

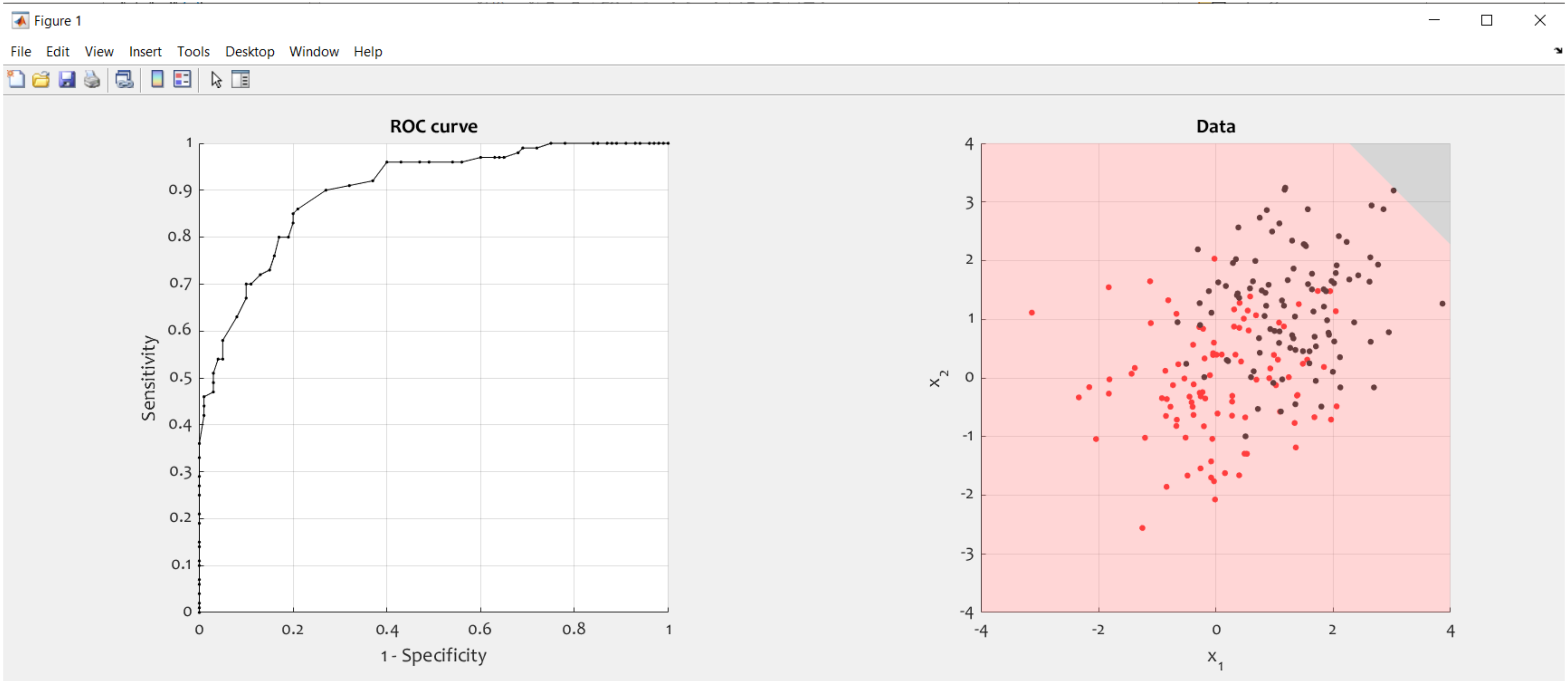
Specificity = 13/21 = 61.9%

1-Specificity = **38.1%**

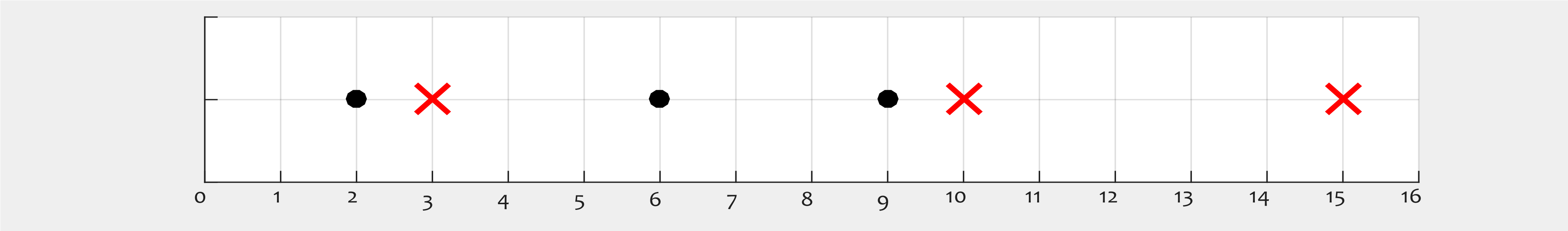
Q2. Plot the operational point for the classifier D, defined by the confusion matrix below

	Positive +	Negative -
Positive +	8	2
Negative -	100	900

ROC curves – MATLAB illustration



Q3. Build a ROC curve for a threshold classifier for the following one-dimensional classification problem (red crosses are class 'positive'):



1. ~~(If the points are given in random order, sort x and re-arrange the labels to match the sorted objects.)~~
2. Suppose that there is a threshold between every pair of consecutive objects. Starting with a threshold on the left of the smallest x, calculate the sensitivity and specificity assuming that all points above the threshold are labelled as class positive (This gives sensitivity 1 and specificity 0).
3. Repeat for all possible thresholds moving to the right, one object at a time. The obtained pairs (sens, spec) will define the points of the ROC curve.
4. Plot sensitivity (y-axis) versus 1-specificity (x-axis). Voila!

Answers to some questions:

Q1.

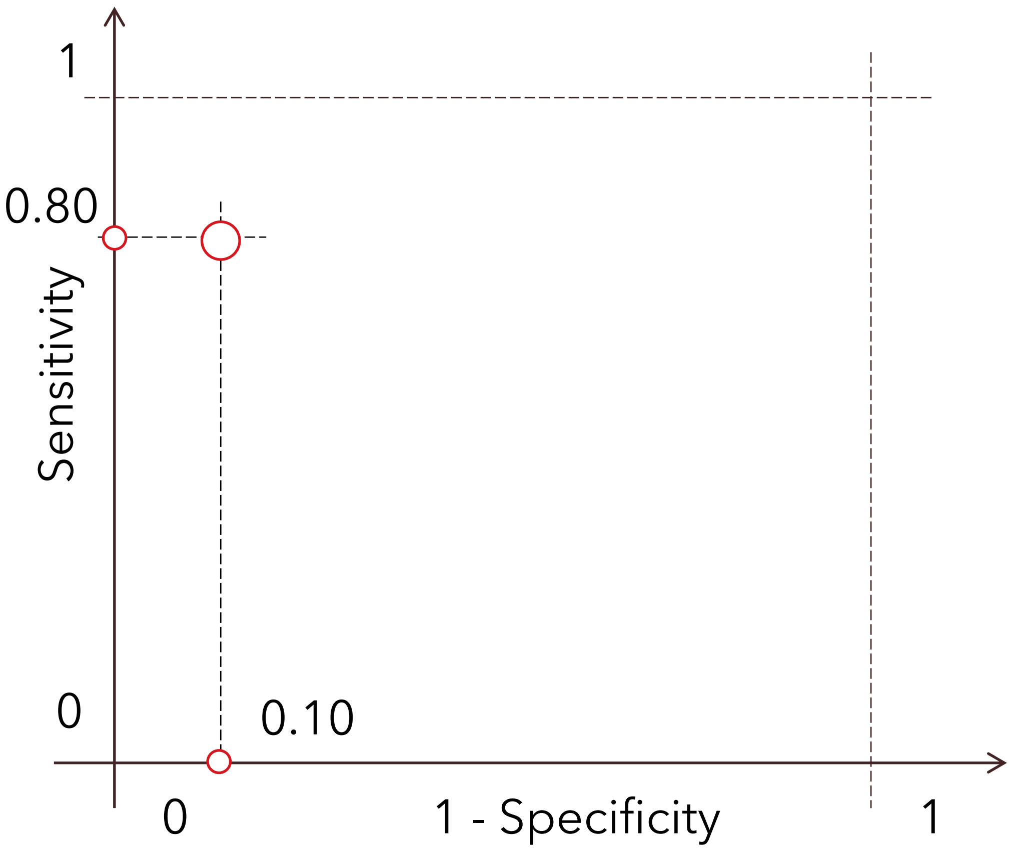
In each row, there are $118 - 25 + 1 = 94$ squares.

The number of rows with squares is $79 - 25 + 1 = 55$ squares.

The total number of squares is therefore $94 \times 55 = 5170.$

The number of negatives is 5170 - 6 = 5164 squares.

Q2 Plot the operational point for the classifier D, defined by the confusion matrix below



	Positive +	Negative -
Positive +	8	2
Negative -	100	900

Sens = 8/(8+2) = 0.8

Spec = 900/(100+900) = 0.9

1-Spec = 1 - 0.9 = 0.1

Q3.