Index Class

1. Classification problems
2. Types of Errors
3. Metrics of classification problems

Classification problems

Classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs.

The prediction outcome is normally in the form of class probability.

- 1 the event is predicted to happen
- 0 the event is not predicted to happen

Metrics - Types of errors

- True positives: predicted = 1, real = 1
- True negative: predicted = 0, real = 0
- False positive: predicted = 1, real = 0
- False negative: predicted = 0, real = 1

Metrics - Types of errors

		Predicted	
		Negative **(N)** **-**	Positive **(P)** **+**
Actual	Negative **-**	True Negative **(TN)**	**False Positive (FP) Type I Error**
	Positive **+**	**False Negative (FN) Type II Error**	True Positive **(TP)**

Metrics - Confusion Matrix

n=165	**Predicted: NO**	**Predicted: YES**	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Metrics - Confusion Matrix

Un estudio investigó 1250 hombres, con sospecha de cáncer de próstata, a los cuales se les realizó una biopsia, 920 dieron positivo a la biopsia y 330 dieron negativo, al finalizar el seguimiento del estudio se confirmó que de los que dieron positivo realmente 750 hombres tenían cáncer de próstata y la prueba tuvo 40 resultados falsos negativos.

Metrics - Types of errors

$$\text{Accuracy (ACC)} = \frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$$

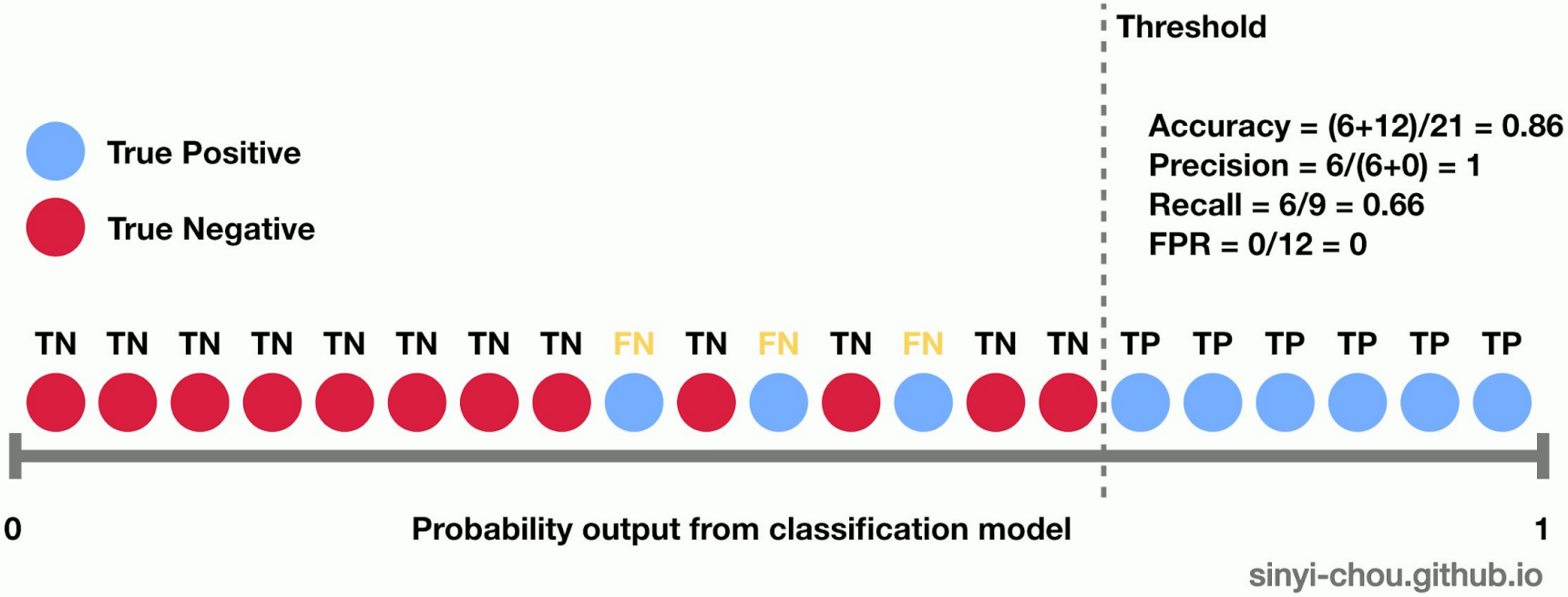
True positive rate (TPR), Recall, Sensitivity, probability of detection

$$= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$$

Specificity (SPC), Selectivity, True negative rate (TNR)

$$= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$$

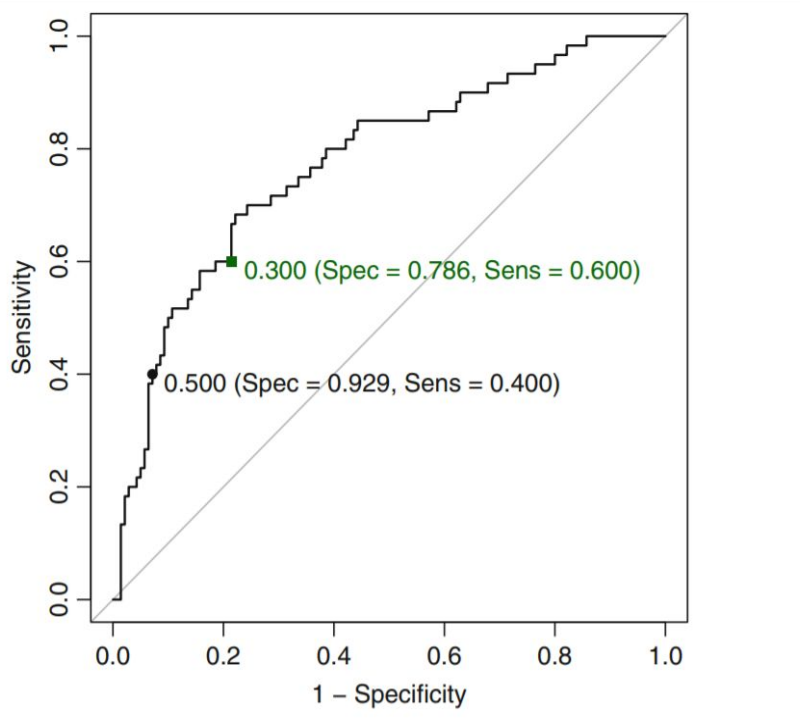
Threshold change metrics



Metrics - ROC Curve

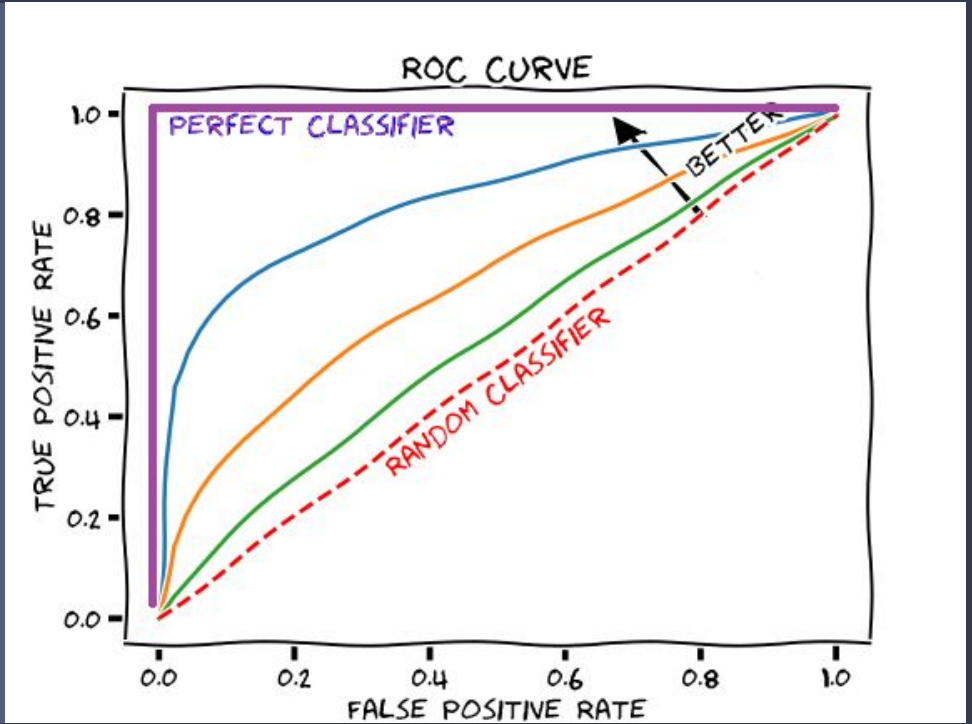
A receiver operating characteristic curve, or ROC curve, is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination **threshold** is varied.

1-Specificity = False Positive Rate
Sensitivity = True Positive Rate

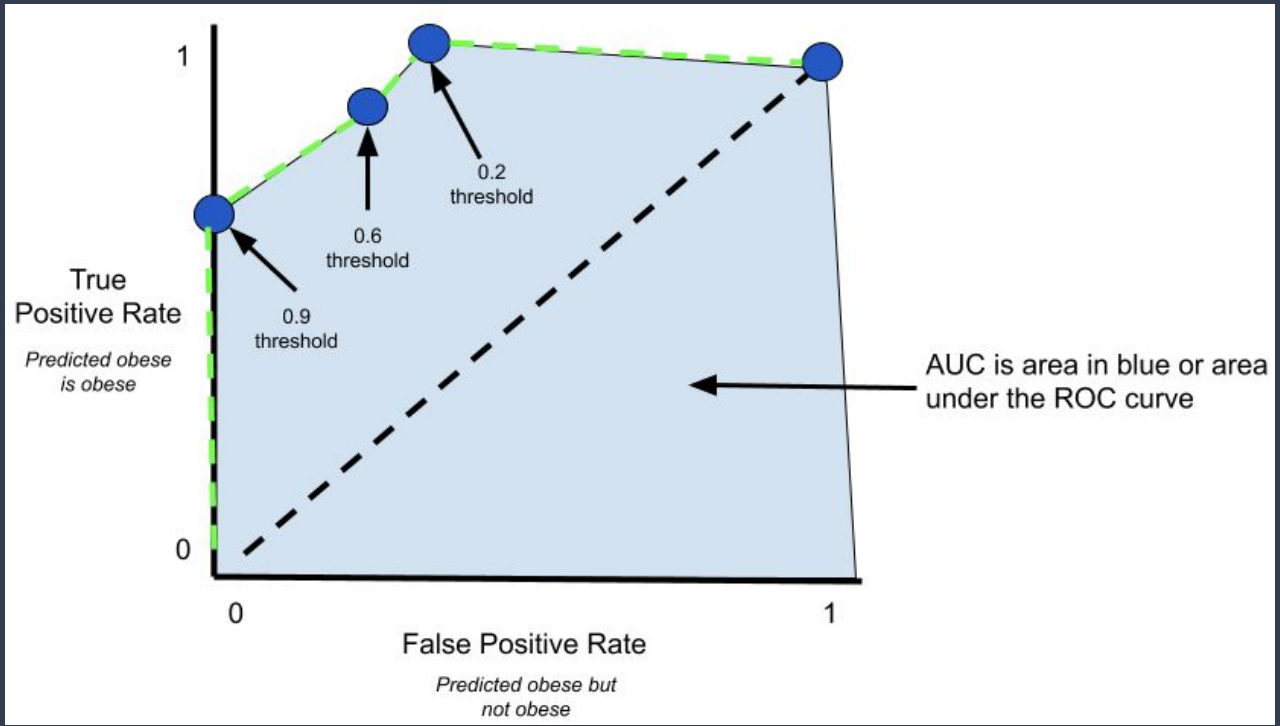


Metrics - AUC

The AUC represents a model's ability to discriminate between positive and negative classes.
1 is the best result
0.5 is as good as random



Metrics - AUC


1
0
0
1
True
Positive Rate
Predicted obese
is obese
False Positive Rate
Predicted obese but
not obese
0.9
threshold
0.6
threshold
0.2
threshold
AUC is area in blue or area
under the ROC curve