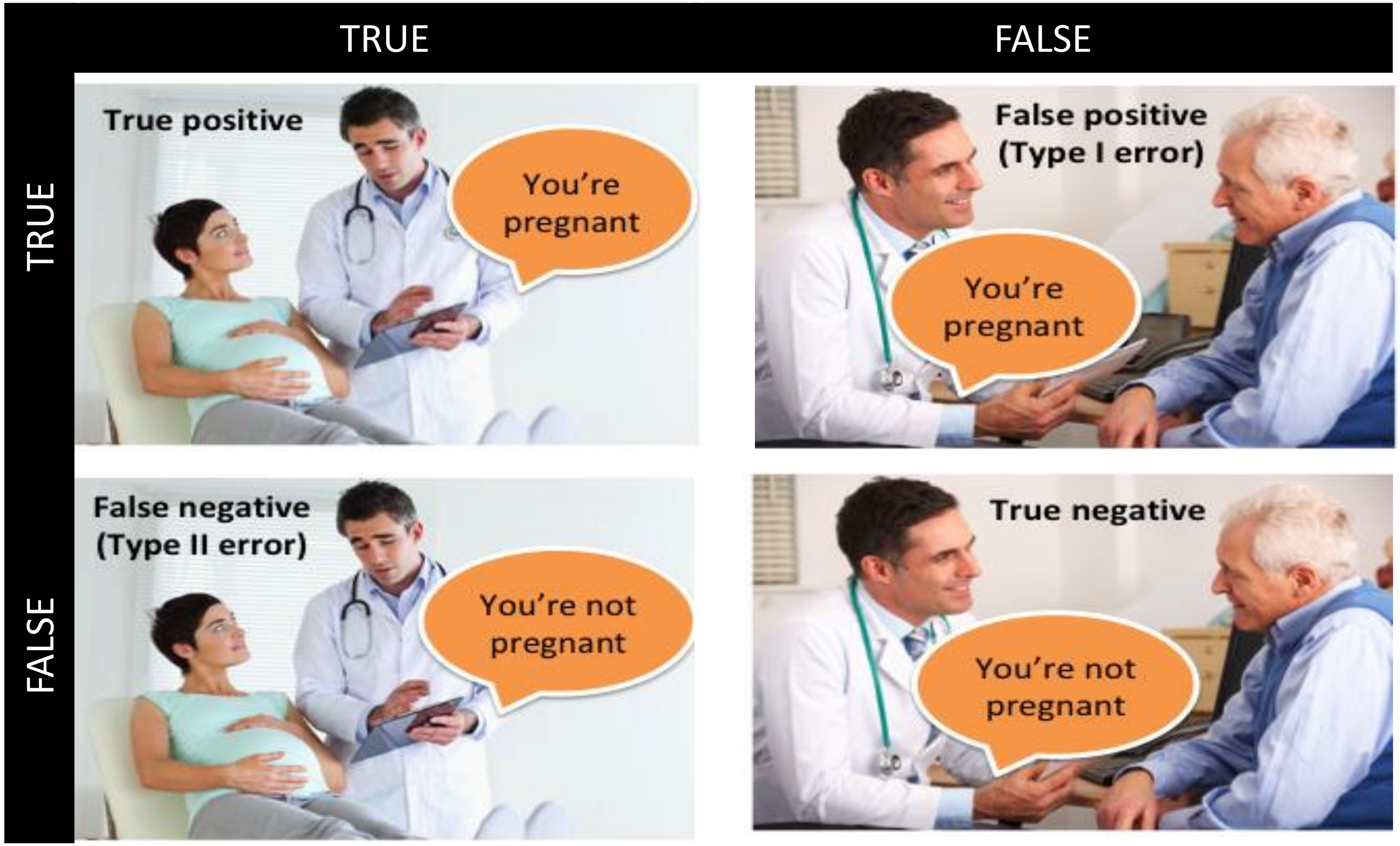
Evaluation of ML Models

- Confusion Matrix
- Accuracy
- Precision & Recall
- F1-Score
- Harmonic Mean
- Specificity & Sensitivity
- AUC & ROC Curve

Confusion Matrix



- TP: Algo predicts Yes, and the actual value is Yes!
- FP: Algo predicts Yes, but the actual value is No!
- FN: Algo predicts No, and the actual value is Yes!
- TN: Algo predicts No, and the actual value is No!

Confusion Matrix

- It is an error matrix with the following form:

	Predicted (False)	Predicted (True)
Actual (False)	True Negative (TN)	False Positive (FP)
Actual (True)	False Negative (FN)	True Positive (TP)

- True positive (TP): the number of <u>correctly</u> predicted samples.
- True negative (TN): the number of <u>correctly</u> predicted samples, as not the considered class.
  – An instance is classified as negative, and it is negative.
- False positive (FP): the number of samples <u>incorrectly</u> predicted as the considered class.
- False negative (FN): the number of samples <u>incorrectly</u> not predicted as the considered class.

Confusion Matrix

Hyp.		Reference: Positive	Reference: Negative
	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Various performance measures:

- True positive rate (hit rate, recall, sensitivity): $\frac{\#TP}{\#TP+\#FN}$
- False positive rate (false alarm rate, fall-out): $\frac{\#FP}{\#FP+\#TN}$
- Positive predictive value (precision): $\frac{\#TP}{\#TP+\#FP}$
- Negative predictive value: $\frac{\#TN}{\#TN+\#FN}$
- True negative rate (specificity): $\frac{\#TN}{\#FP+\#TN} = 1 -$ false positive rate

Confusion Matrix

Metric	Definition
Accuracy	The proportion of correct predictions.
Precision	The proportion of positive predictions that are actually correct.
Recall	The proportion of actual positives that are correctly identified.
F1-score	A harmonic mean of precision and recall.

How can we Test the Performance of our Classifier?

In pattern recognition, the performance of a classifier can be evaluated using various metrics that measure its accuracy and effectiveness in recognizing and classifying different patterns. Some common metrics-

1. **Confusion matrix:** A confusion matrix is a table that summarizes the number of correct and incorrect predictions made by a classifier. It provides a breakdown of how many patterns were classified into each category or class, as well as how many patterns were misclassified.
2. **Accuracy:** Accuracy is a measure of how well a classifier performs in terms of the percentage of correctly classified patterns. It is calculated as the ratio of the number of correctly classified patterns to the total number of patterns.
3. **Precision and Recall:** Precision measures the proportion of correctly classified patterns among those that were predicted to belong to a particular class. Recall measures the proportion of correctly classified patterns among those that actually belong to a particular class.
4. **F1-score:** The F1-score is a measure of the balance between precision and recall. It is calculated as the harmonic mean of precision and recall and provides a single score that summarizes the overall performance of a classifier.
5. **ROC curve and AUC:** The Receiver Operating Characteristic (ROC) curve represents the trade-off between the true positive rate and false positive rate for different classification thresholds.

These metrics can be used to evaluate the performance of a classifier on a test dataset, a separate dataset not used during training. We can select the best classifier for a particular pattern recognition task by comparing the performance of different classifiers using these metrics.

Accuracy / Recognition Rate

- Accuracy = (TP + TN) / (TP + FP + TN + FN)

- Condition positive (P).
  - The number of real positive cases in the data.

- Condition negative (N).
  - The number of real negative cases in the data.

Precision & Recall

- Consider a binary classifier with the positive and negative class.
- Precision is the ratio between the true positive samples and all the positive samples. It can be defined as:
- $Precision = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP)+False\ Positive\ (FP)}$
- Recall is the measure of the correctly classified true positive samples. It can be defined as:
- $Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP)+False\ Negative\ (FN)}$
- Related measures:
  - Sensitivity is the recall. It corresponds to the proportion of the positive class that got correctly classified.
  - Specificity the proportion of the negative class of the negative class that go correctly classified. It is given by: $Specificity\ = \frac{True\ Negative\ (TN)}{True\ Negavtive\ (TN)+False\ Positive\ (FP)}$

F1 Score & ROC

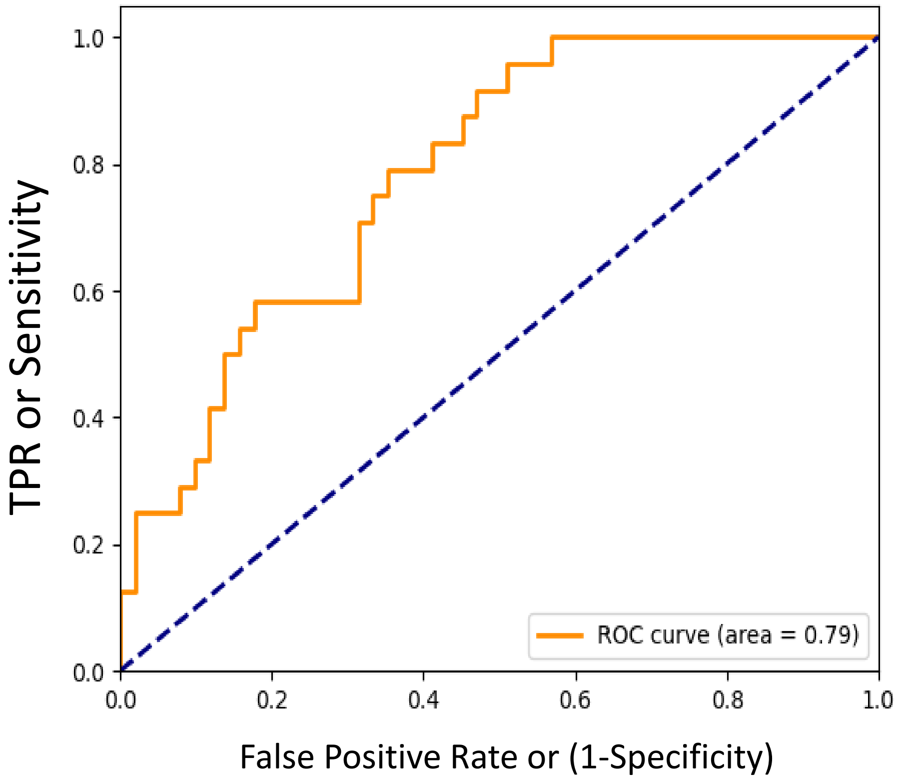
- F1-Score: it is the harmonic mean of the precision and recall. It is given by:

$$F_1 = 2 * \frac{Precision \ * Recall}{Precision + Recall}$$

- The Receiver Operating Curve (ROC) is a graph showing the classification performance of a model at all classification thresholds. This curve plots two parameters:
  - True Positive Rate (Recall).
  - False Positive Rate (False alarm).

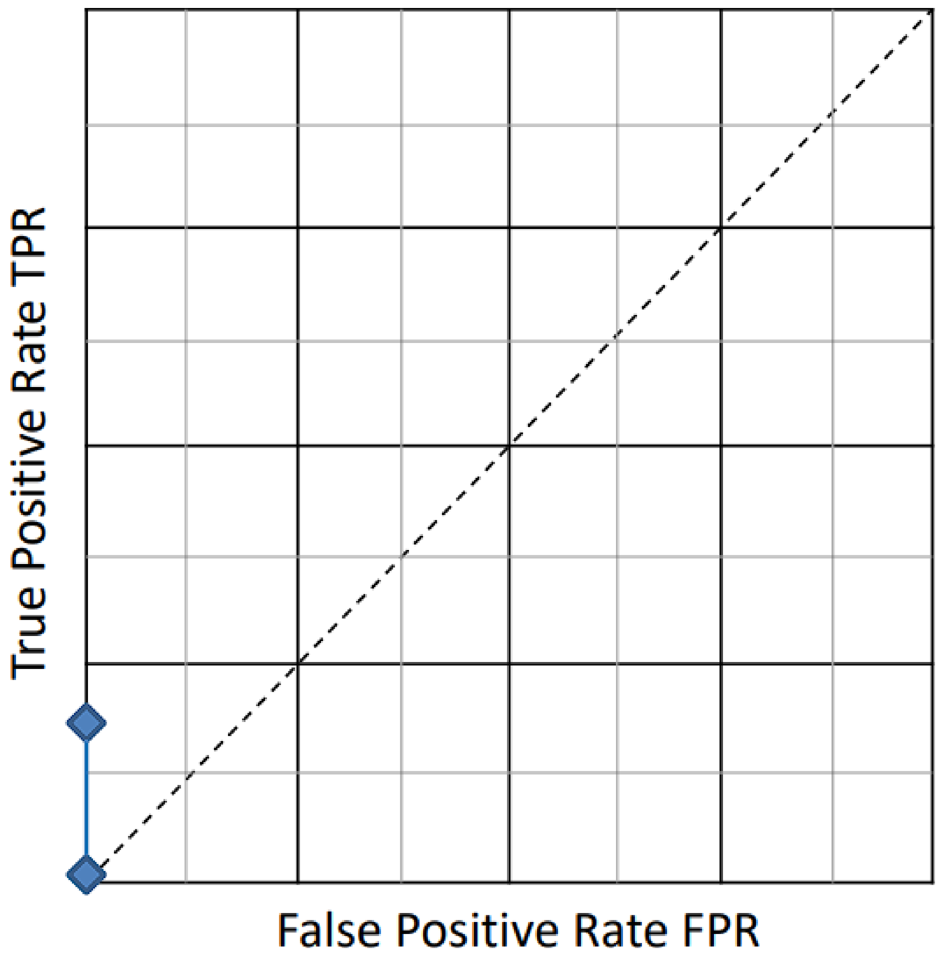
ROC & AUC Curve

Receiver Operating Characteristic (ROC): Since, TPR is equivalent to Sensitivity and FPR is equal to 1 − specificity, the ROC graph is sometimes called the sensitivity vs (1 − specificity) plot.



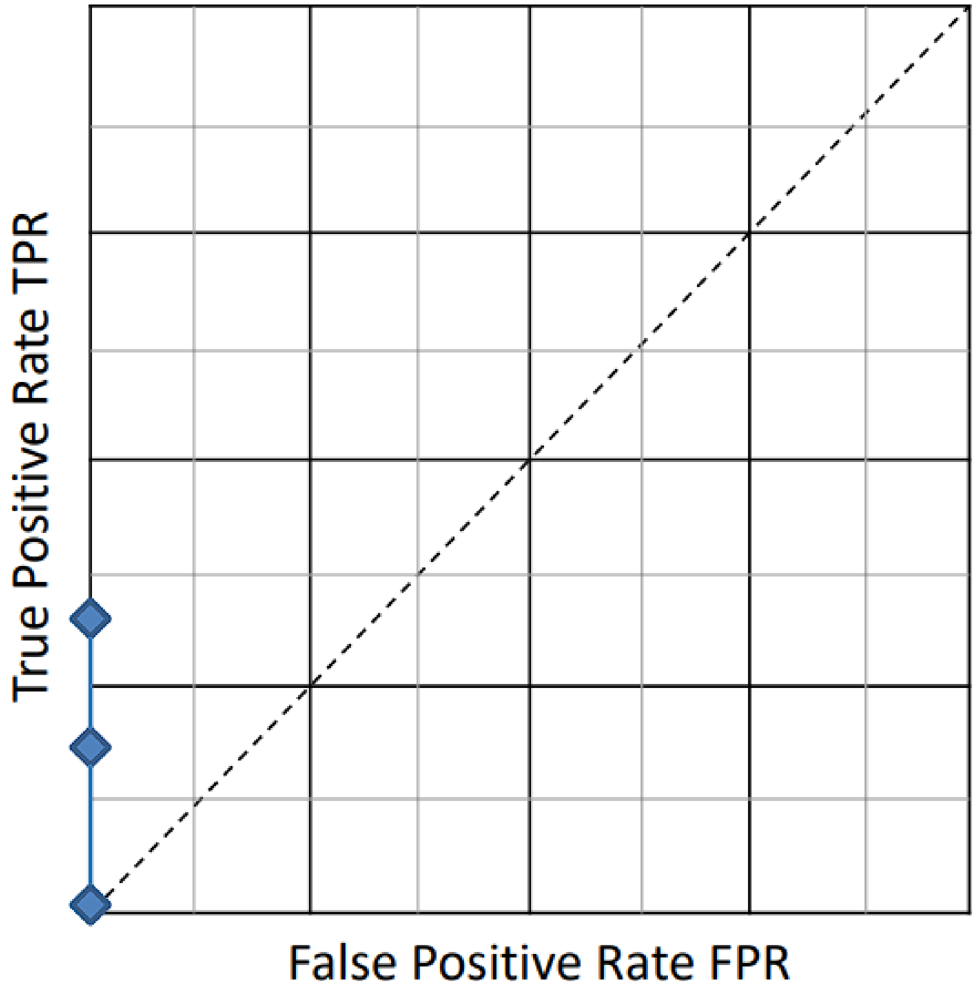
Example of Receiver Operating Curve (ROC)

	Predict Correct	Score
1	**Yes**	**0.95**
2	Yes	0.86
3	Yes	0.69
4	No	0.65
5	Yes	0.59
6	No	0.52
7	No	0.39
8	No	0.28
9	Yes	0.15
10	No	0.06



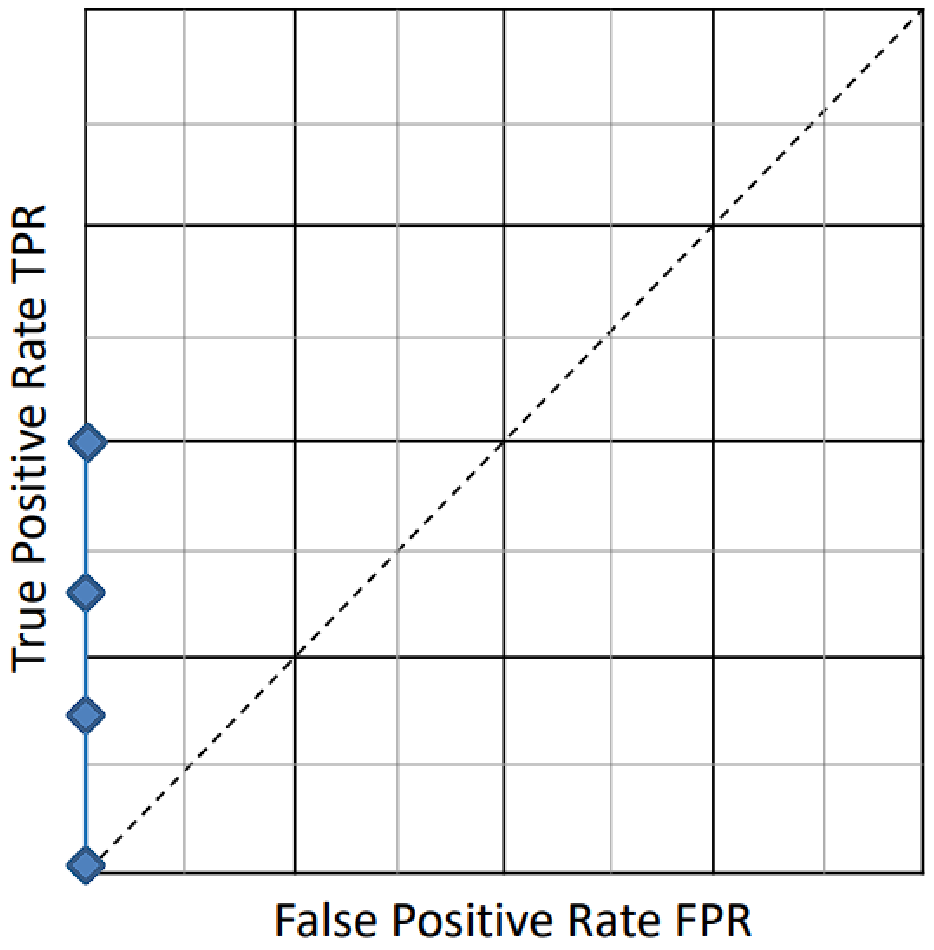
Example of Receiver Operating Curve (ROC)

	Predict Correct	Score
1	**Yes**	**0.95**
2	**Yes**	**0.86**
3	Yes	0.69
4	No	0.65
5	Yes	0.59
6	No	0.52
7	No	0.39
8	No	0.28
9	Yes	0.15
10	No	0.06



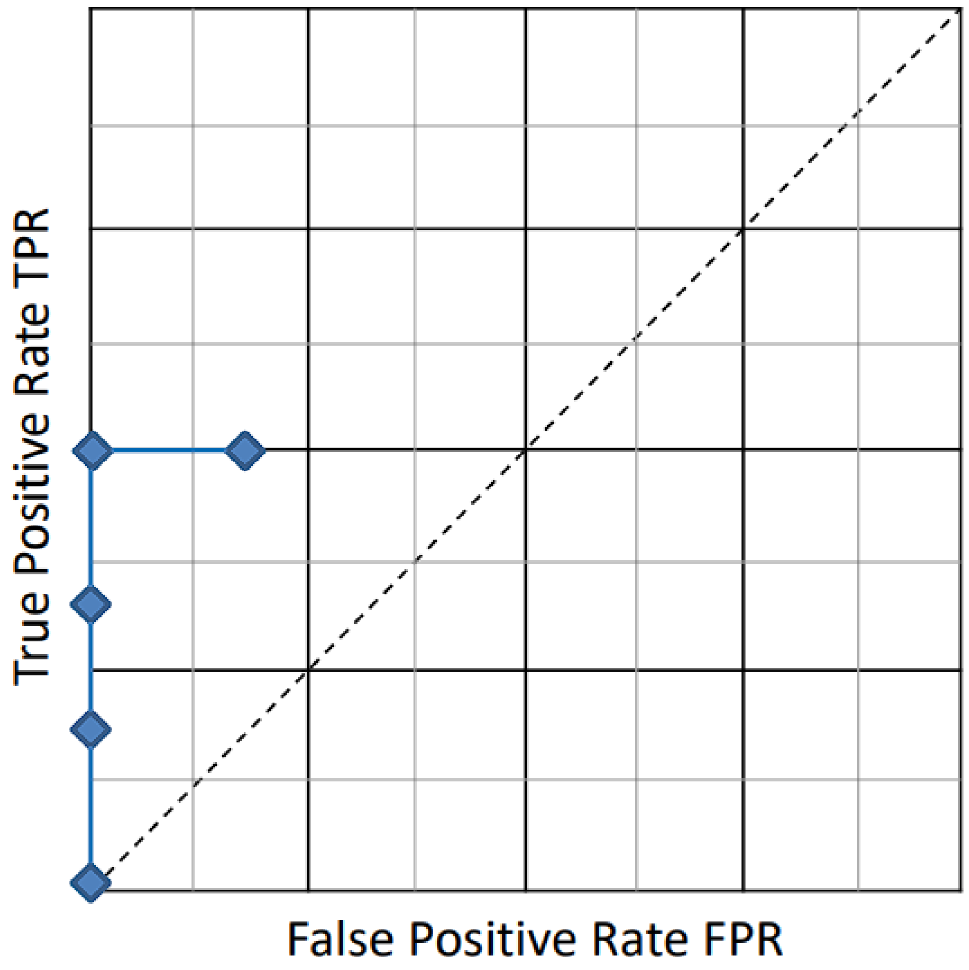
Example of Receiver Operating Curve (ROC)

	Predict Correct	Score
1	**Yes**	**0.95**
2	**Yes**	**0.86**
3	**Yes**	**0.69**
4	No	0.65
5	Yes	0.59
6	No	0.52
7	No	0.39
8	No	0.28
9	Yes	0.15
10	No	0.06



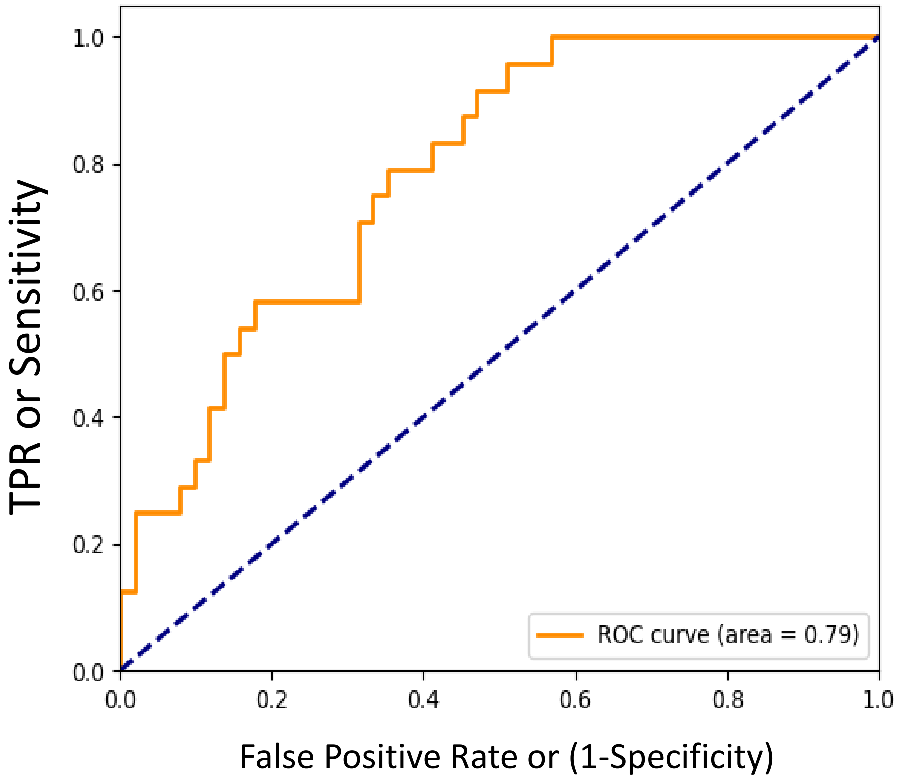
Example of Receiver Operating Curve (ROC)

	Predict Correct	Score
1	**Yes**	**0.95**
2	**Yes**	**0.86**
3	**Yes**	**0.69**
4	**No**	**0.65**
5	Yes	0.59
6	No	0.52
7	No	0.39
8	No	0.28
9	Yes	0.15
10	No	0.06



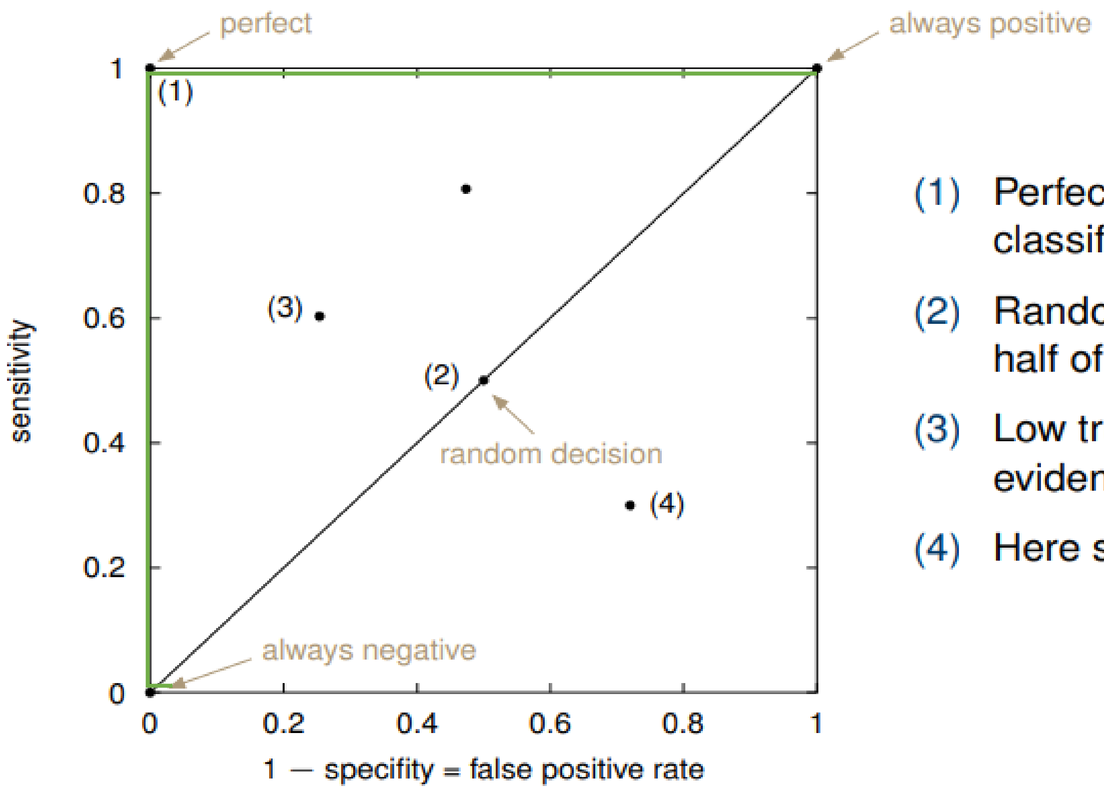
ROC & AUC Curve

Receiver Operating Characteristic (ROC): Since, TPR is equivalent to Sensitivity and FPR is equal to 1 − specificity, the ROC graph is sometimes called the sensitivity vs (1 − specificity) plot.



ROC & AUC Curve

Receiver Operating Characteristic (ROC): Since, TPR is equivalent to Sensitivity and FPR is equal to 1 − specificity, the ROC graph is sometimes called the sensitivity vs (1 − specificity) plot.



(1) Perfect classifier: no false positives, all negatives are classified as negatives

(2) Random decision: half of positives are classified correctly, half of negatives are classified correctly

(3) Low true positive rate, but lower false positive rate; strong evidence for positive classification

(4) Here something goes really wrong!

ROC & AUC Curve

The Area Under the Curve (AUC) measures a classifier's ability to differentiate between classes and is used as a summary of the ROC curve. The higher the AUC, the better the performance of the model at distinguishing between the positive and negative classes.

ROC & AUC Curve

- **In general,** an AUC of 0.5 suggests no discrimination (like the ability to diagnose patients with and without the disease or condition based on the test)
- 0.7 to 0.8 is considered **acceptable**
- 0.8 to 0.9 is considered **excellent**, and
- more than 0.9 is considered **outstanding.**