Introduction to Data Science

(Lecture 19)

Dr. Mohammad Pourhomayoun
Assistant Professor
Computer Science Department
California State University, Los Angeles

Model Evaluation and Error Measurements

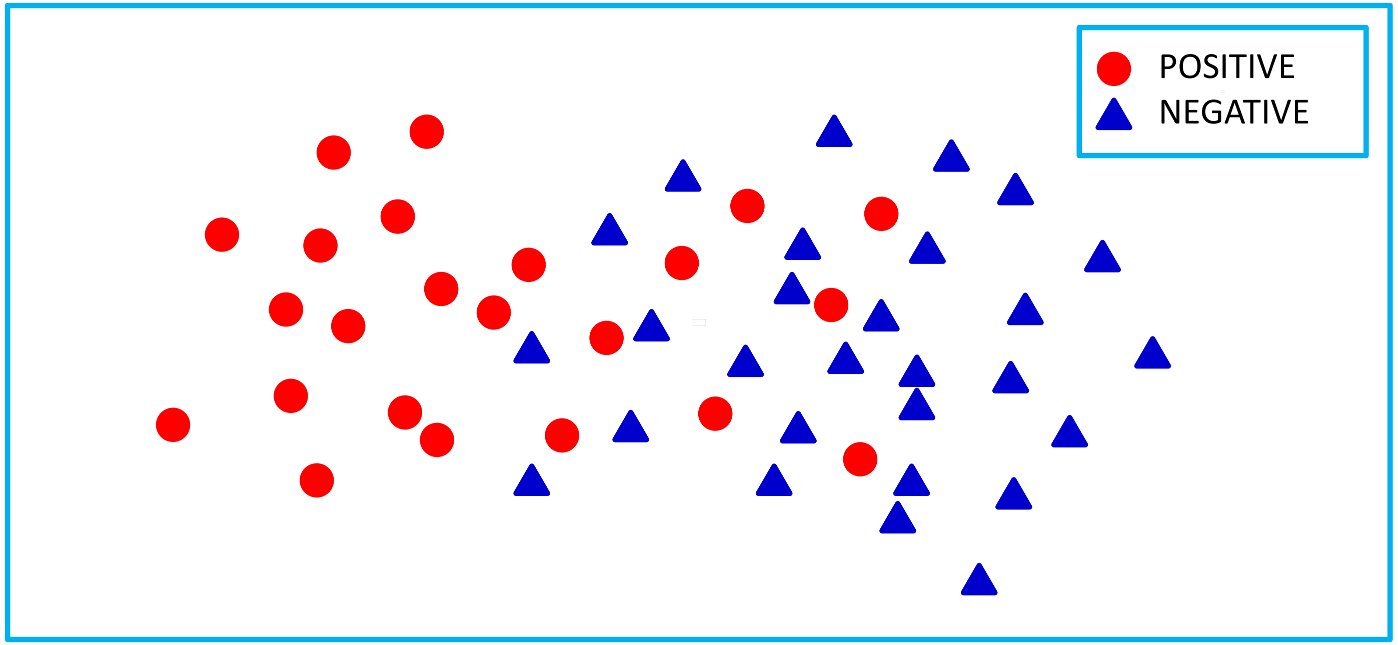
Positive and Negative Labels

- In binary Classification, we have two labels.
- One of the labels that usually shows "**happening of an Event**" is called **Positive** Label.
- The other one that usually shows "**not happening of the Event**" is called **Negative** Label.

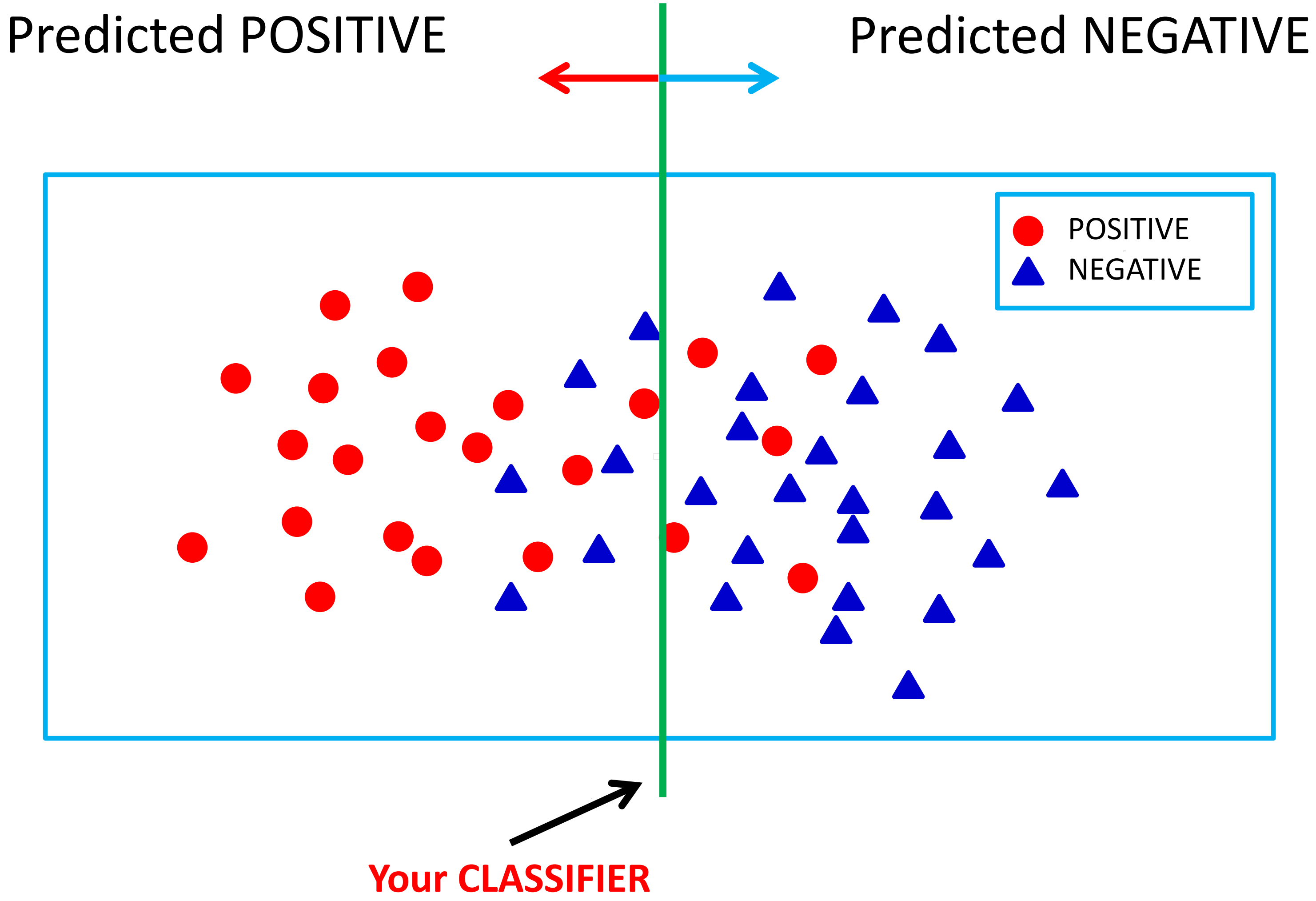
- **Example**:
  - Positive: Rainy, Negative: Sunny (Not-Rainy)
  - Positive: Spam, Negative: Not-Spam
  - Positive: Cancer, Negative: Not-Cancer
  - Positive: Heart Attack, Negative: Not- Heart Attack
  - …

Prediction Accuracy

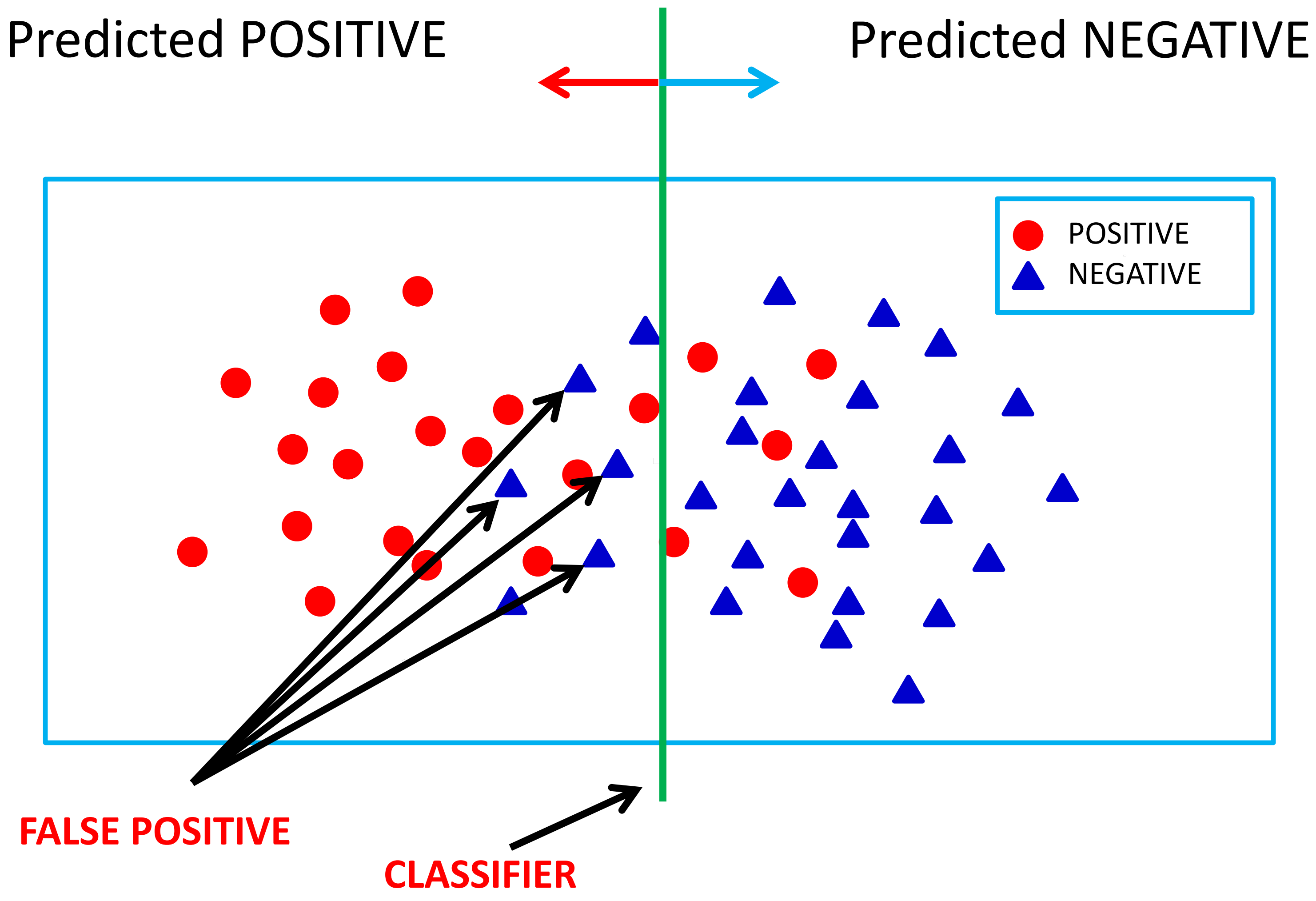
The original labeled data:



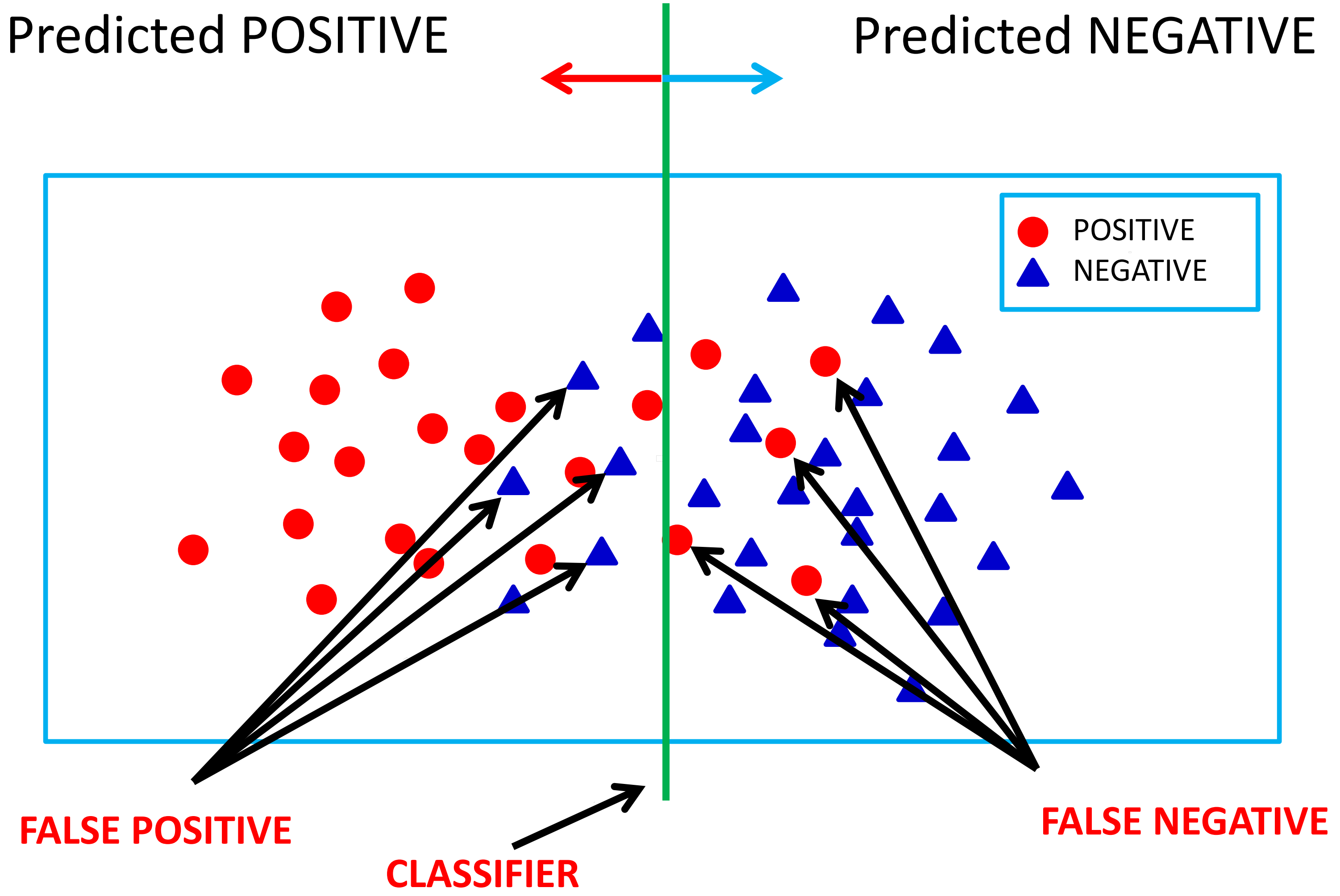
Prediction Accuracy



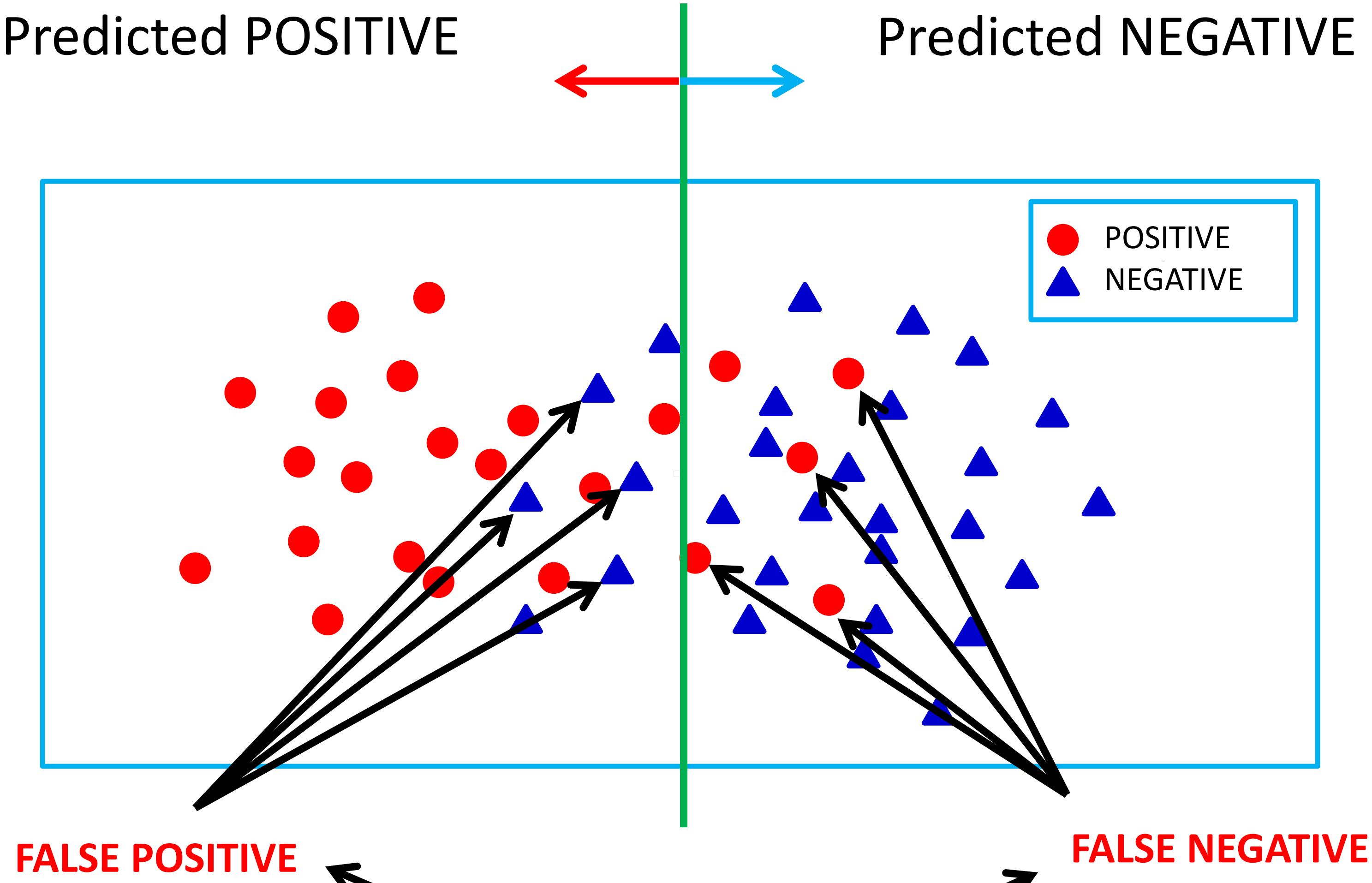
Prediction Accuracy



Prediction Accuracy



Prediction Accuracy



BOTH COSTLY

Which one is worse?

Two Types of Error

- **FP = False Positive (False Alarm):**
  - It was actually negative, BUT we incorrectly predicted as positive.
- **FN = False Negative (Miss):**
  - It was actually positive, BUT we incorrectly predicted as negative.

Definitions

- **TP = True Positive:**
  - It was actually positive, and we predicted as positive.
- **TN = True Negative:**
  - It was actually negative, and we predicted as negative.
- **FP = False Positive (False Alarm):**
  - It was actually negative, BUT we incorrectly predicted as positive.
- **FN = False Negative (Miss):**
  - It was actually positive, BUT we incorrectly predicted as negative.

Confusion Matrix

		Predicted Label	
		POSITIVE	NEGATIVE
Actual Label	POSITIVE	?	?
	NEGATIVE	?	?

Confusion Matrix

		Predicted Label	
		POSITIVE	*NEGATIVE*
Actual Label	*POSITIVE*	TRUE POSITIVE	FALSE NEGATIVE
	NEGATIVE	FALSE POSITIVE	TRUE NEGATIVE

Accuracy

- So far we have just used **Accuracy** to evaluate a predictive model.
- As we learned before, **Accuracy** is the **percent of correctly classified samples:**

$$\text{Accuracy} = \frac{\text{Number of Correctly Classified Samples}}{\text{Total Number of Samples}}$$

$$= \frac{TP + TN}{TP + TN + FP + FN}$$

What is wrong with Accuracy?

- Accuracy does not care about the likelihood of labels!

- Example: The probability of observing Cancer in regular timely basis Mammography is less than 1%.
  - So, 99% of the times, the answer is No.
  - Rather than designing a Machine Learning algorithm to detect the cancer from the medical image, can I always say NO!!?, and in this case, I still achieve 99% accuracy without even checking the Mammography results!!?

- **So, we may need a better Metric!**

Sensitivity and Specificity

- ***True Positive Rate (TPR)*, also Called *Sensitivity*** is the percent of correct predictions for positive samples.

$$TPR = \frac{\textbf{Number of Correctly Classified Positives}}{\textbf{Total Number of Positives}}$$

$$= \frac{TP}{All\ Positives}$$

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

- ***Sensitivity (TPR)*** tells us how much of the real 'Positive' cases are detected. Or, How well can it **detect** the Events?

Sensitivity and Specificity

- ***True Negative Rate (TNR)***, **also Called** ***Specificity*** is the percent of correct predictions for negative samples.

$$TNR = \frac{\text{Number of Correctly Classified Negatives}}{\text{Total Number of Negatives}}$$

$$= \frac{TN}{All\ Negatives}$$

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

- ***Specificity (TNR)*** tells us how much of the real 'Negative' cases are detected. Or, How well can it **rule out** the Events?

False Positive Rate (FPR)

$$\textbf{FPR} = 1 - \boldsymbol{Specificity}(\boldsymbol{TNR})$$

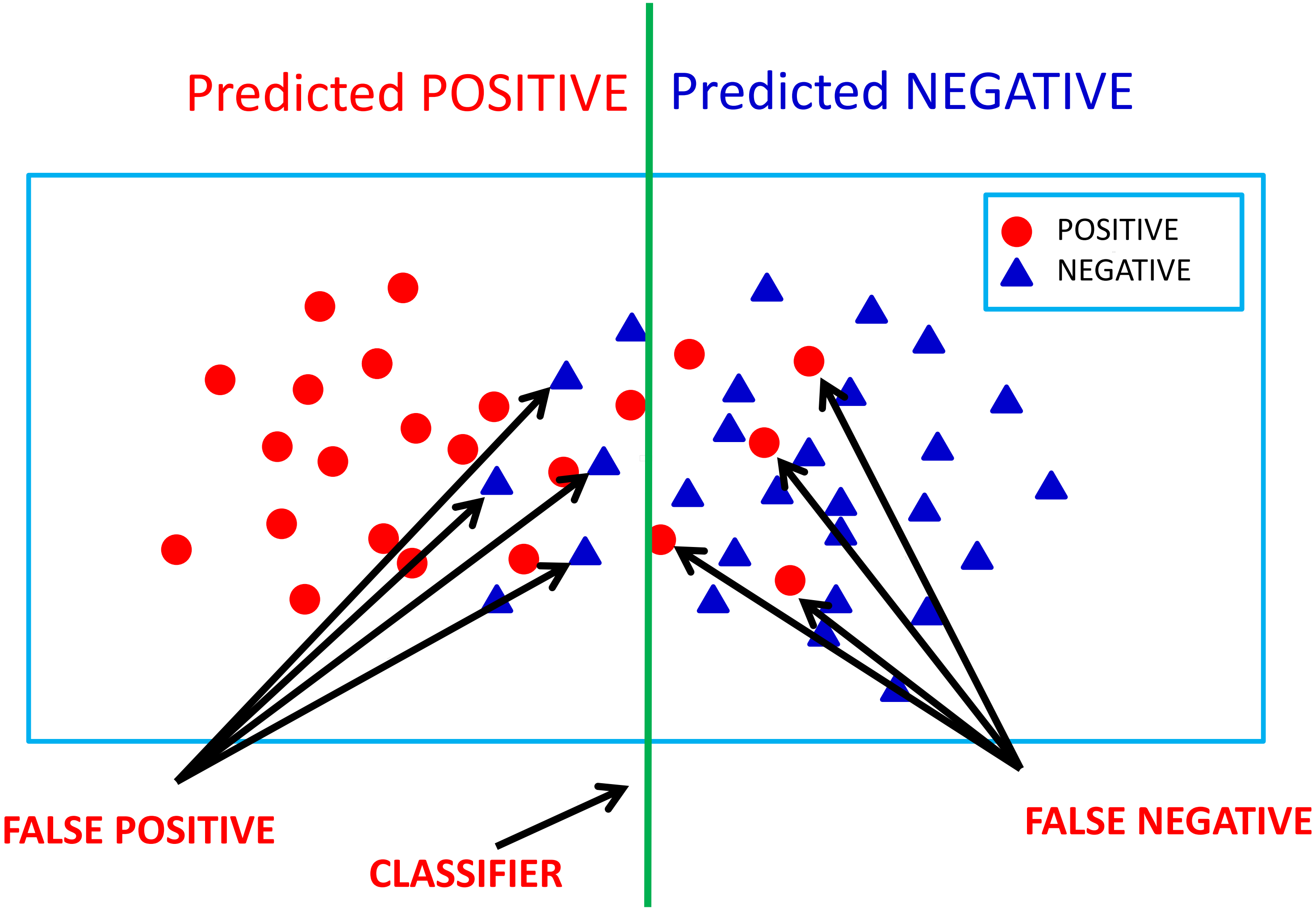
$$= \frac{FP}{All\ Negatives}$$

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

- **False Positive Rate** is also called **False Alarm Rate**.

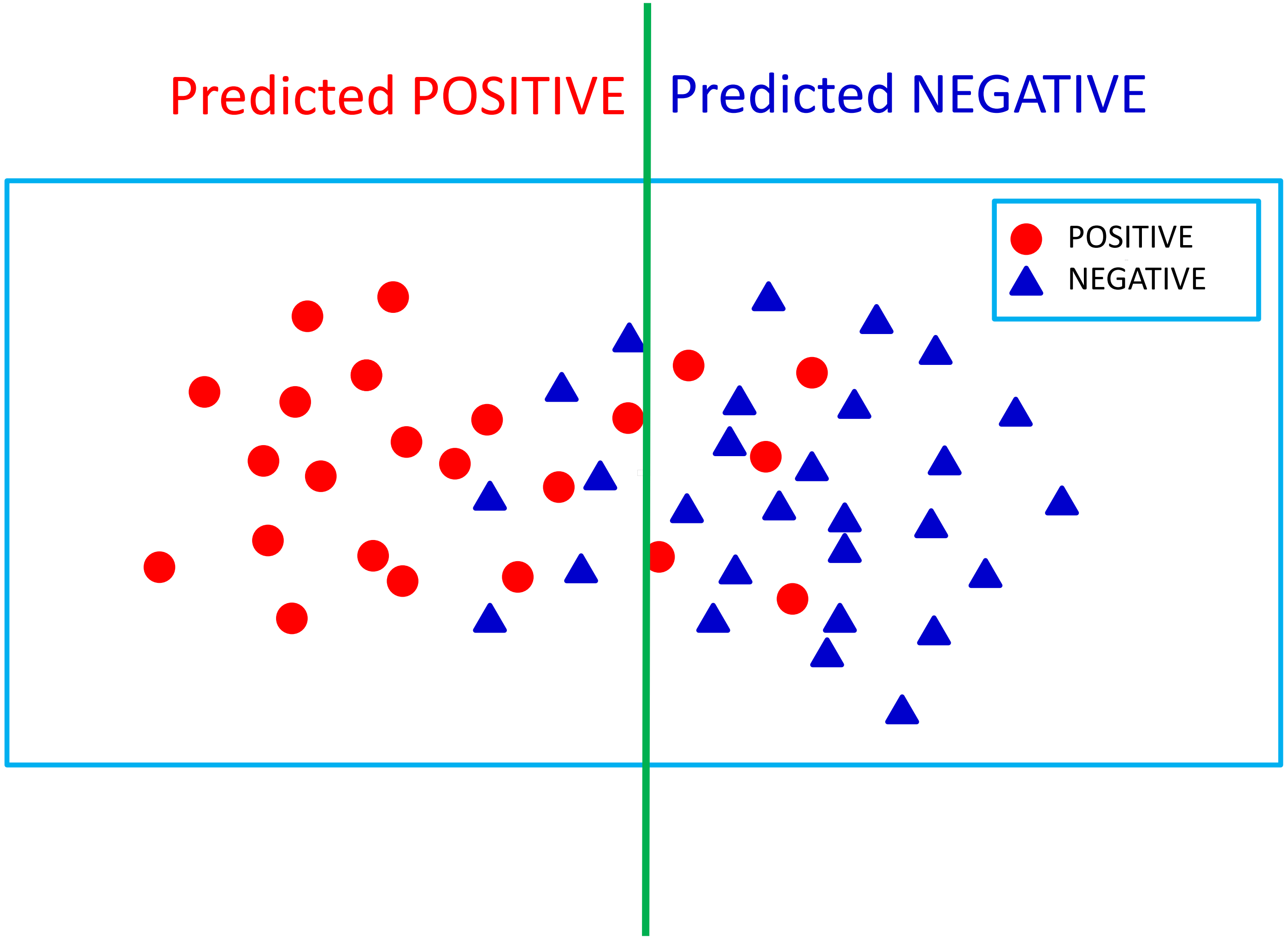
$$\textit{Sensitivity} = \textit{TPR} = \frac{TP}{All\ Positives}$$

$$\textit{Spicificity} = \textit{TNR} = \frac{TN}{All\ Negatives}$$



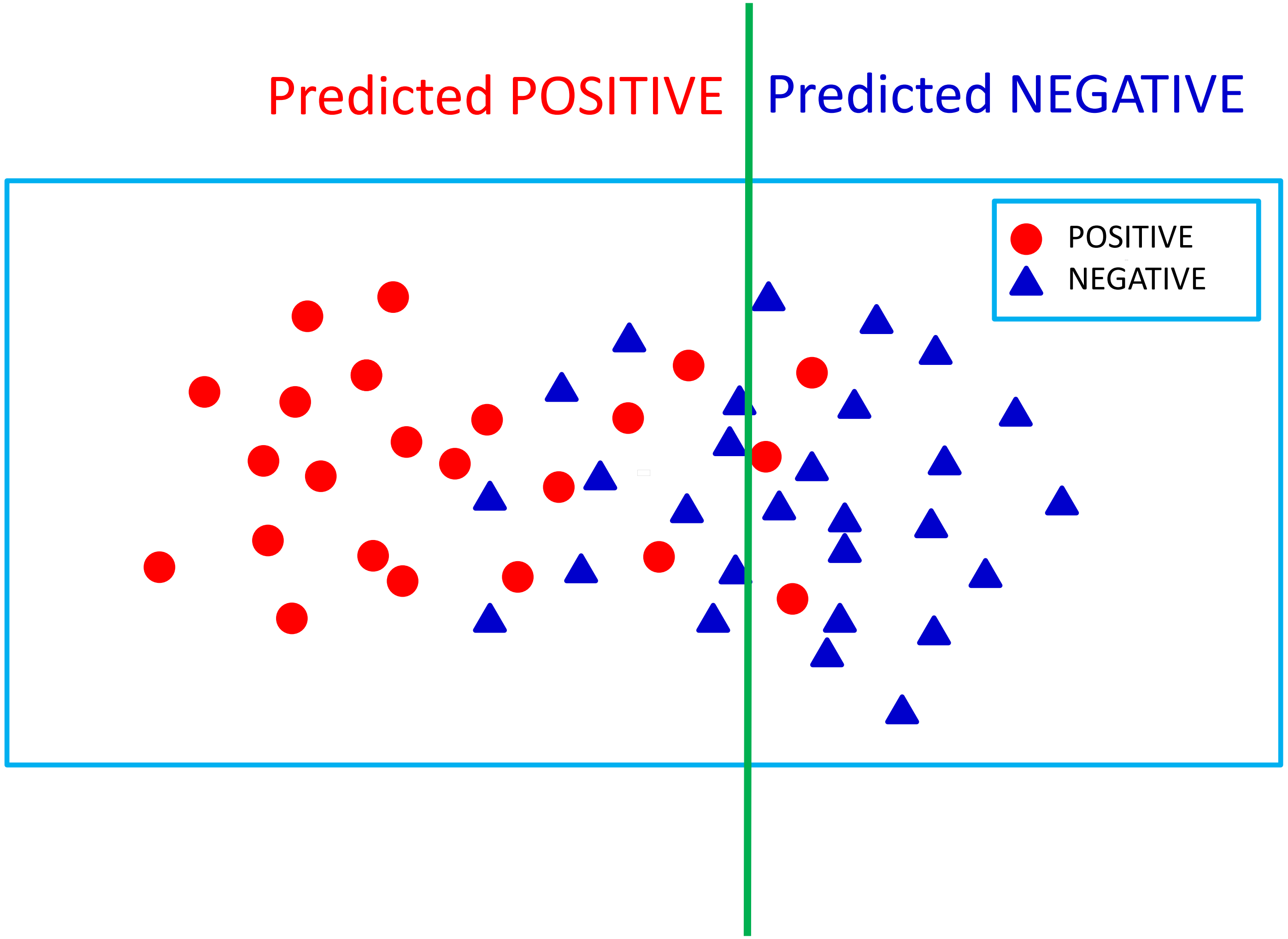
Sensitivity = TPR $= \dfrac{TP}{All\ Positives}$

Spicificity = TNR $= \dfrac{TN}{All\ Negatives}$



$$\uparrow \textbf{\textit{Sensitivity = TPR}} = \frac{TP}{All\ Positives}$$

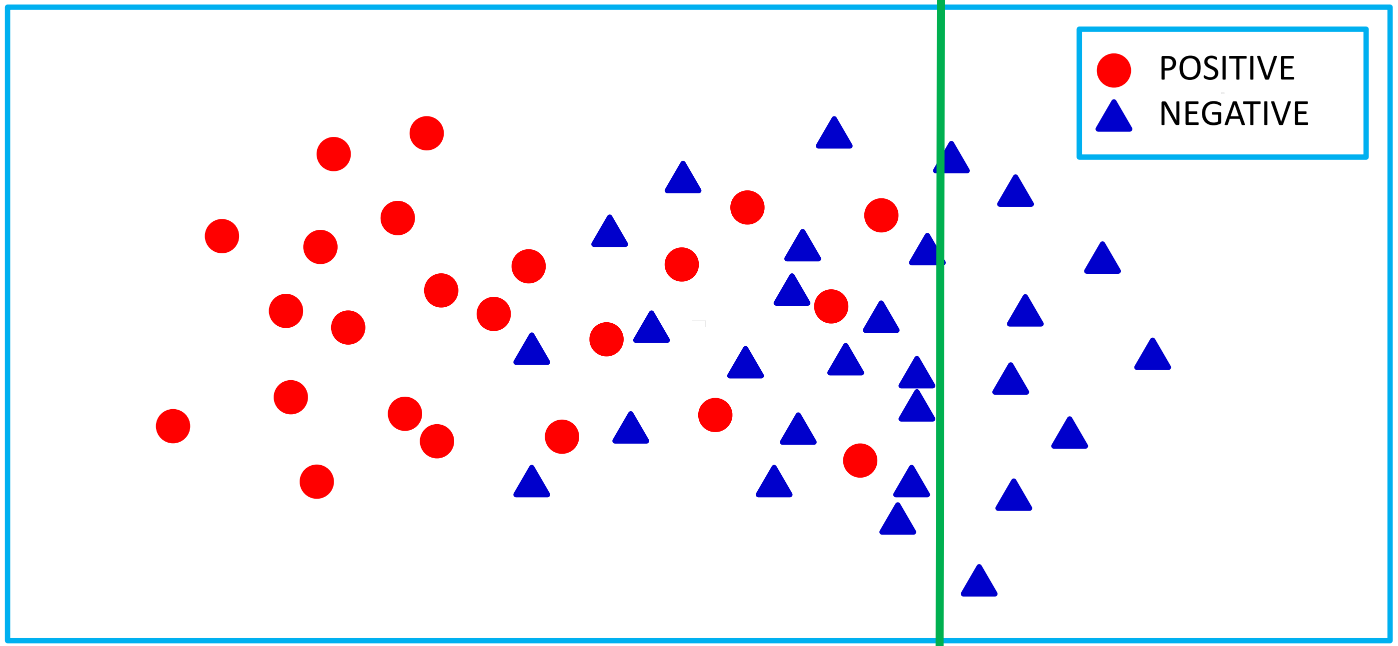
$$\downarrow \textbf{\textit{Spicificity = TNR}} = \frac{TN}{All\ Negatives}$$



$\uparrow$ **Sensitivity = TPR** $= \dfrac{TP}{All\ Positives}$ **= 100%**

$\downarrow$ **Spicificity = TNR** $= \dfrac{TN}{All\ Negatives}$

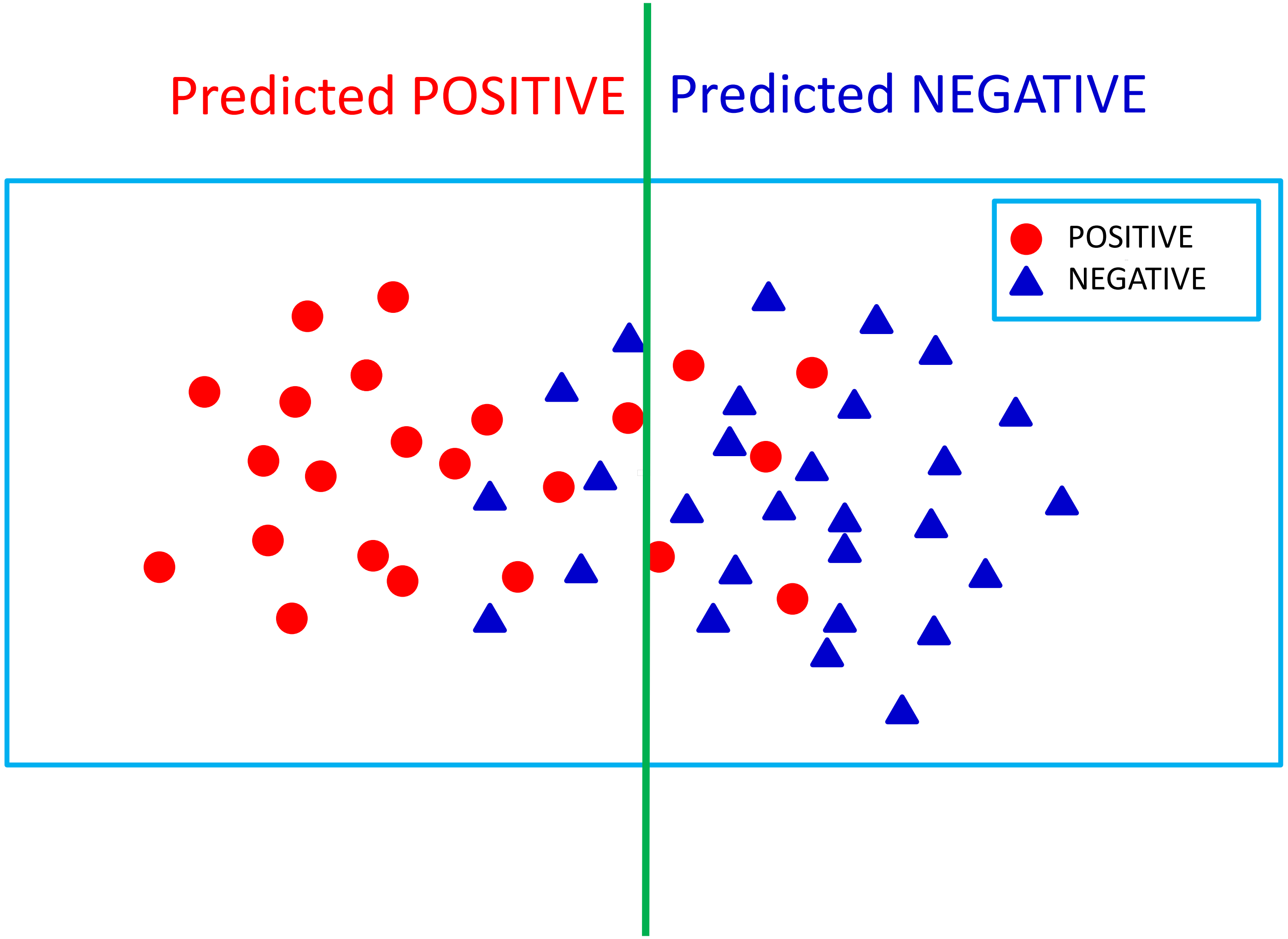
Predicted POSITIVE | Predicted NEGATIVE



100% Sensitivity means: detects *all* Positive Samples but with many false positives (many false alarms).

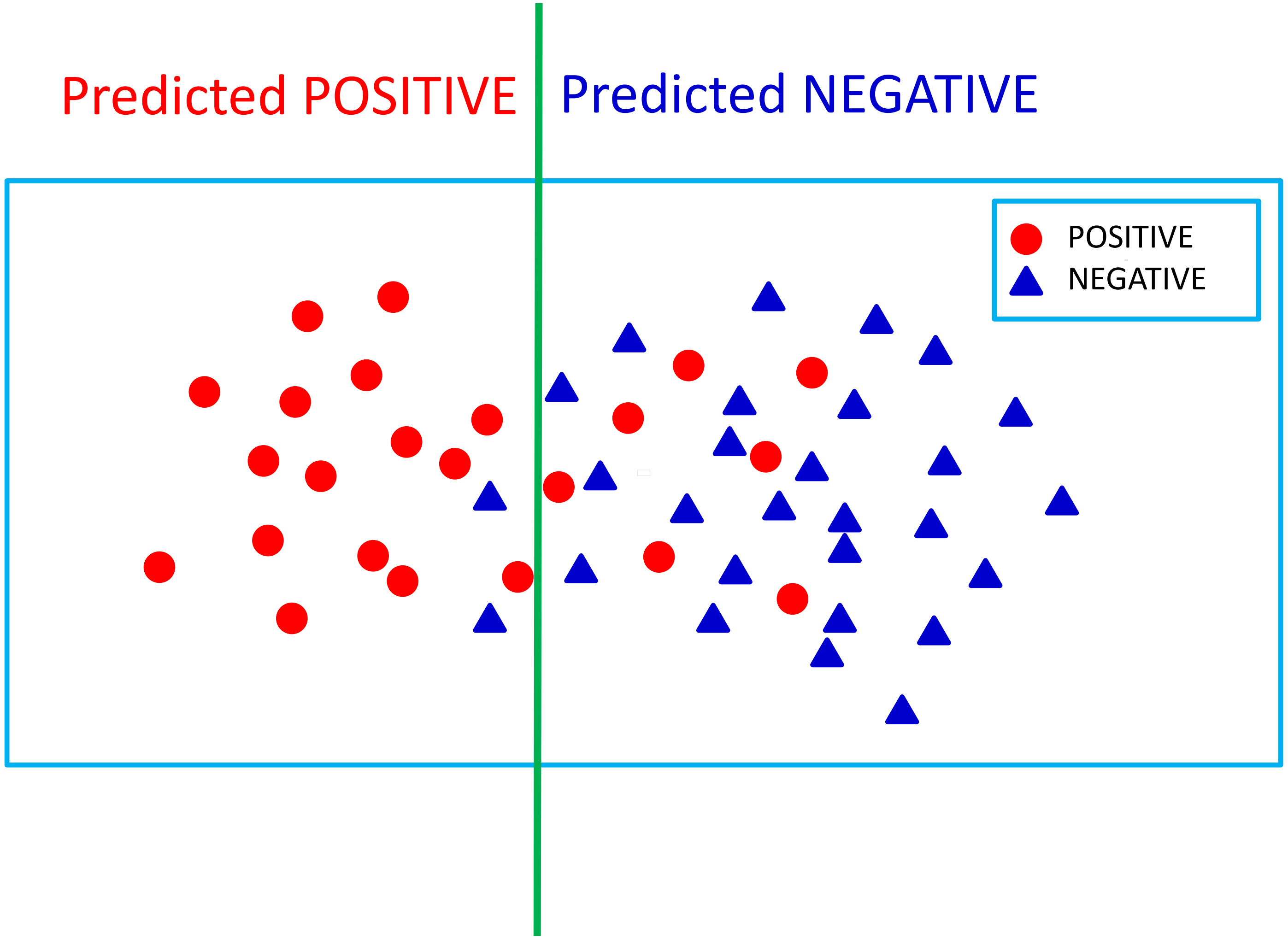
$$\textbf{Sensitivity = TPR} = \frac{TP}{All\ Positives}$$

$$\textbf{Spicificity = TNR} = \frac{TN}{All\ Negatives}$$



$\downarrow$ ***Sensitivity = TPR*** $= \dfrac{TP}{All\ Positives}$

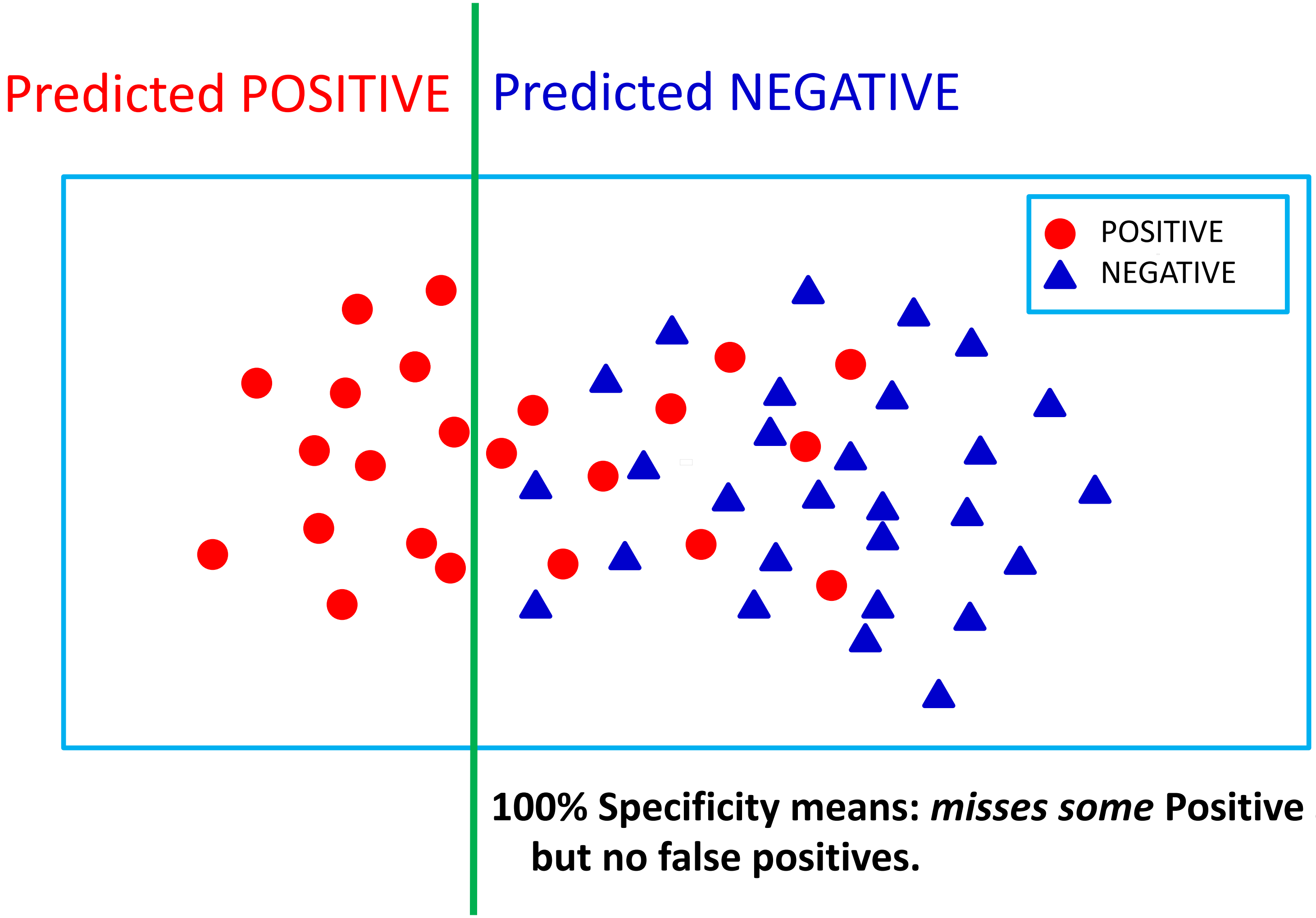
$\uparrow$ ***Spicificity = TNR*** $= \dfrac{TN}{All\ Negatives}$



$\downarrow$ ***Sensitivity = TPR*** $= \dfrac{TP}{All\ Positives}$

$\uparrow$ ***Spicificity = TNR*** $= \dfrac{TN}{All\ Negatives}$

= 100%



100% Specificity means: *misses some* Positive Samples but no false positives.

In Summary

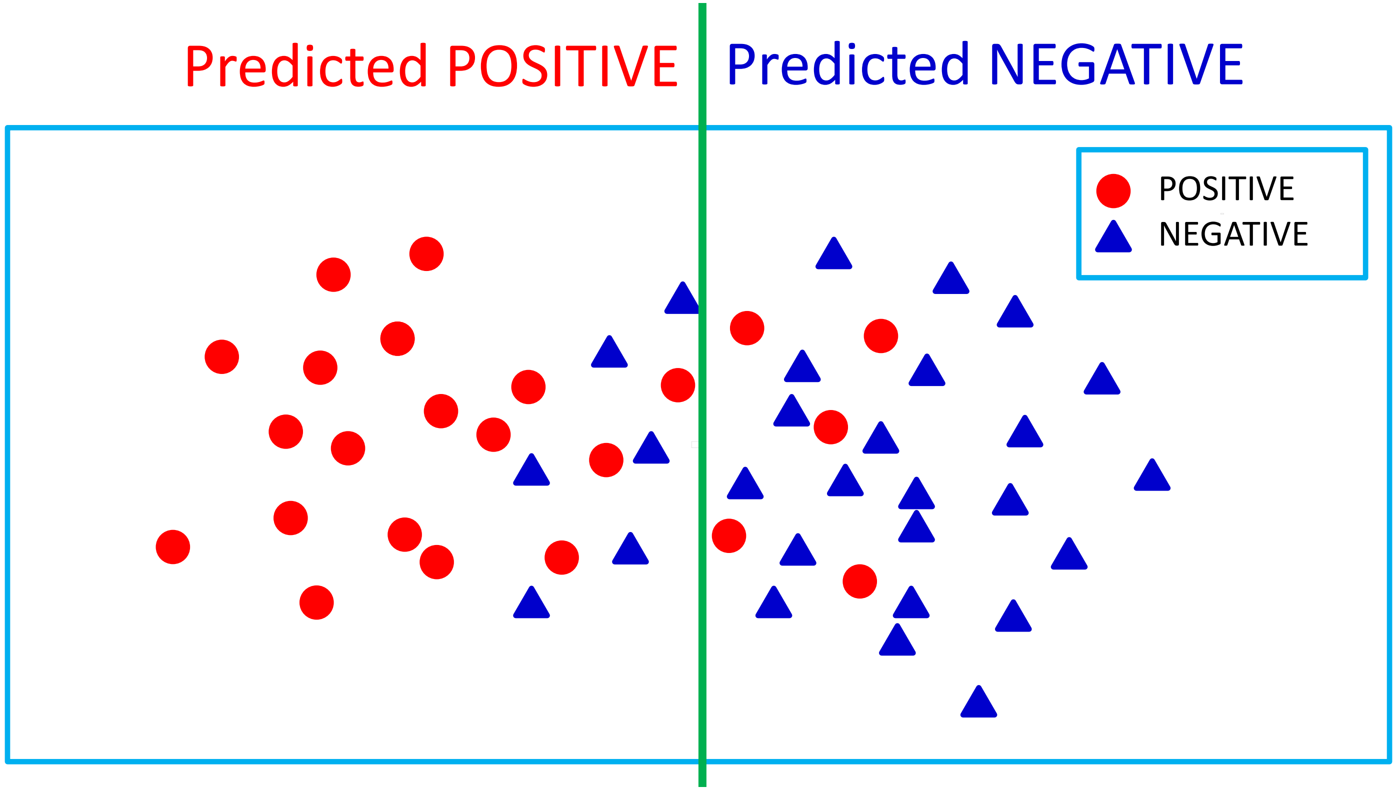
- There is a trade off between TPR (Sensitivity) and TNR (Specificity).
- Depending on the application, we can select a classifier that meets our desired TPR and TNR.
- Since "FPR = 1- TNR", we will have direct relation between TPR and FPR.

ROC Curves

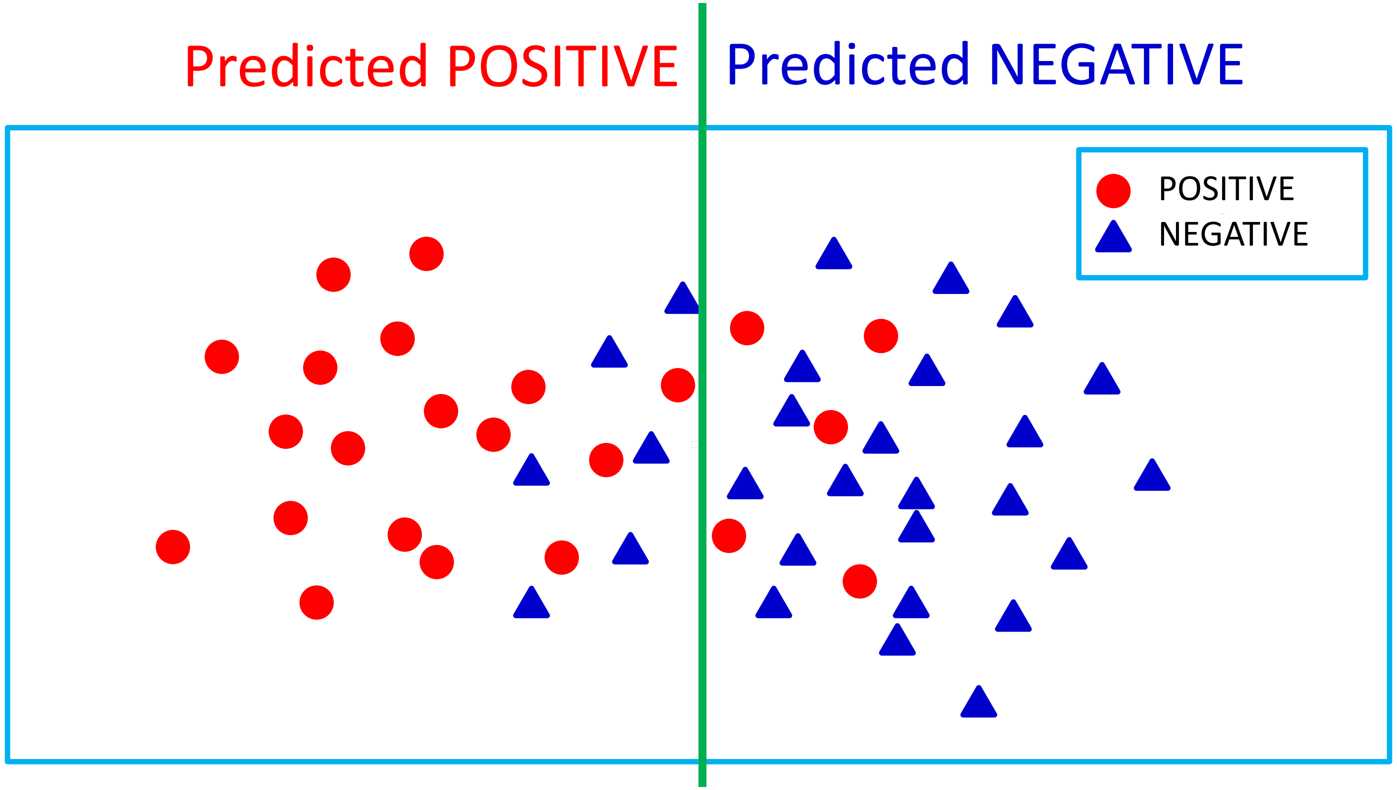
ROC Curves

- ***ROC* = *Receiver Operating Characteristic***
- The ROC curve was first developed by electrical engineers in radio signal detection theory (1940s-1950s).
- Later, it became very popular in medicine, radiology, biometrics, and other applications of machine learning and data science.

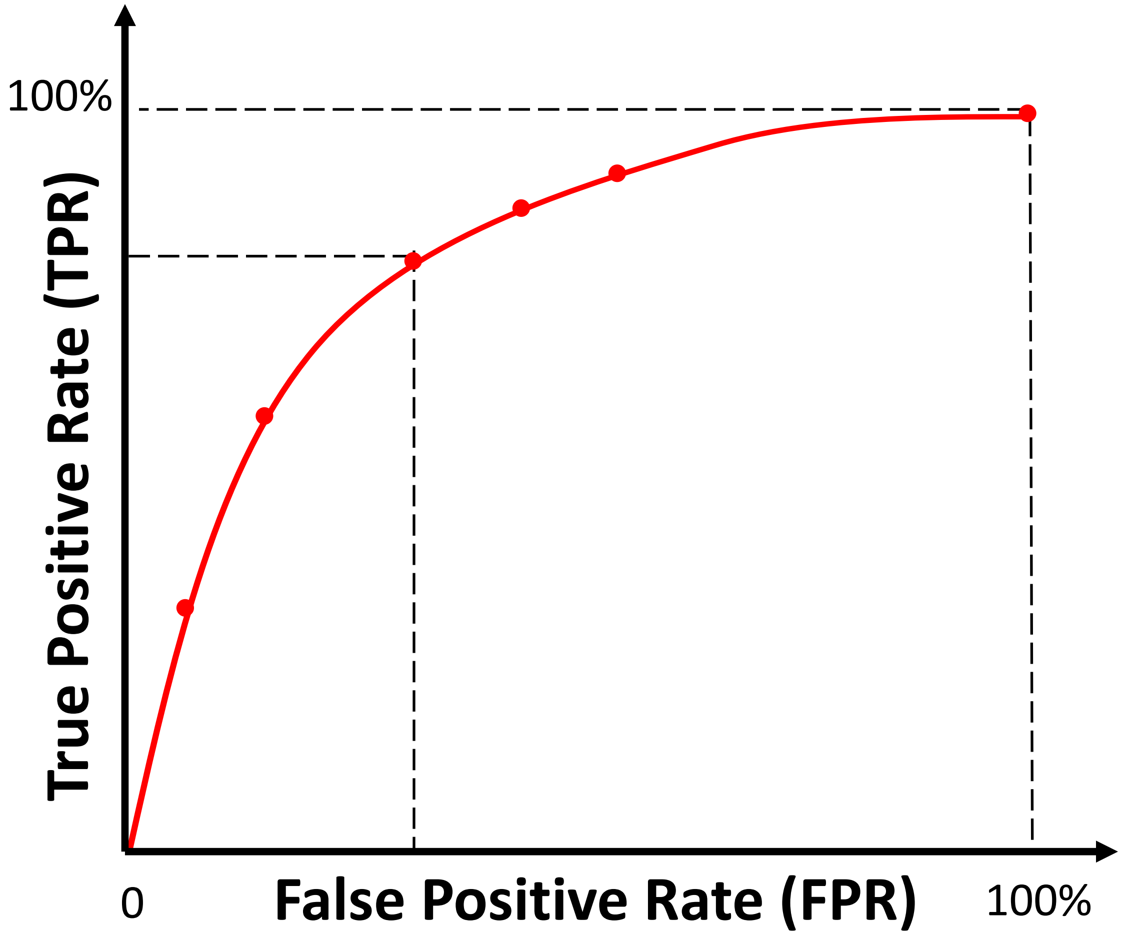
- As we saw, it is possible to change the **TPR** and **TNR** (or **FPR**) by adjusting the classifier.
- For example, in this figure, **TPR** will **increase** by shifting the classifier to the right. The **FPR** (False Alarm Rate) will also **increase** since we generate more false alarms!
- On the other hand, **TPR** will **decrease** by shifting the classifier to the left. The **FPR** (False Alarm Rate) will also **decrease** since we generate less false alarms!

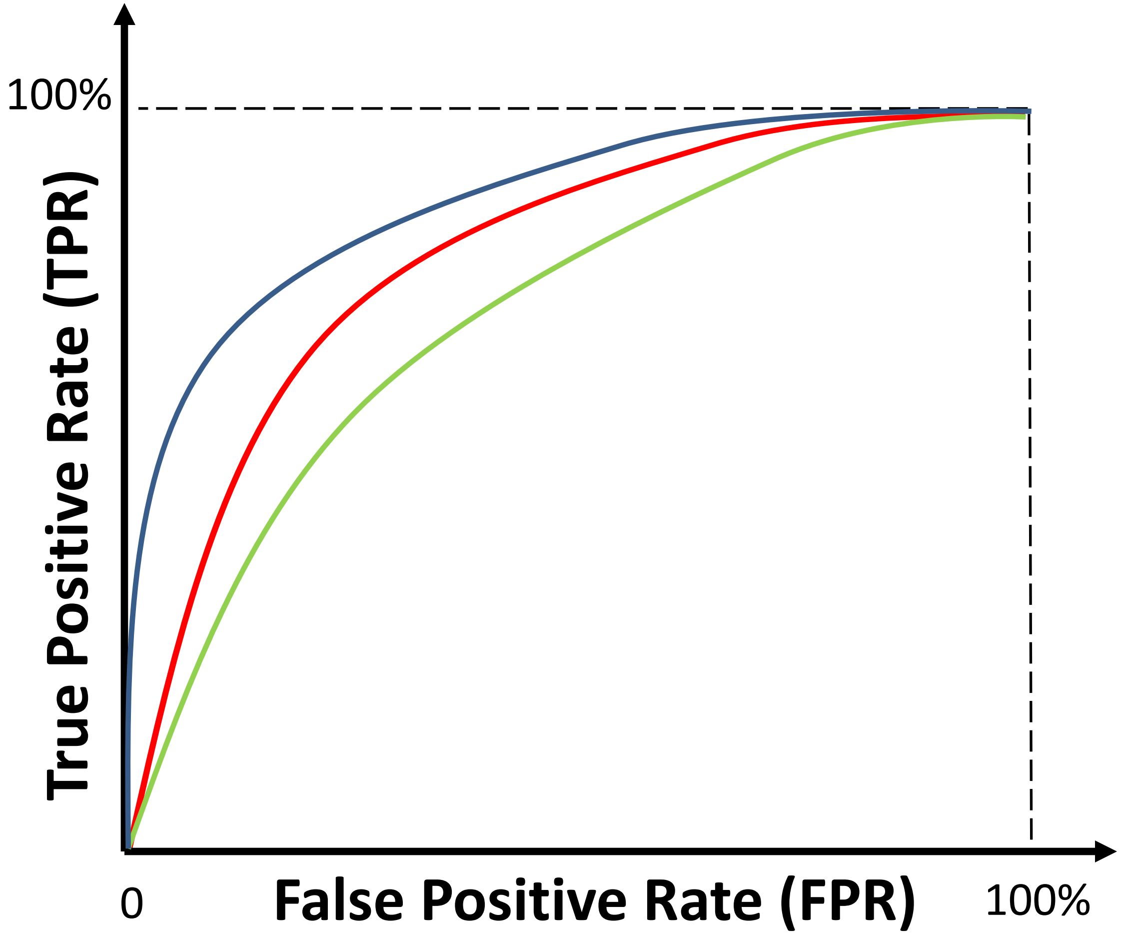


- In this simple example, Let’s move the classifier line back and forth to generate some **(TPR,FPR)** points.
- **ROC is your model curve in TPR vs. FPR plane!**

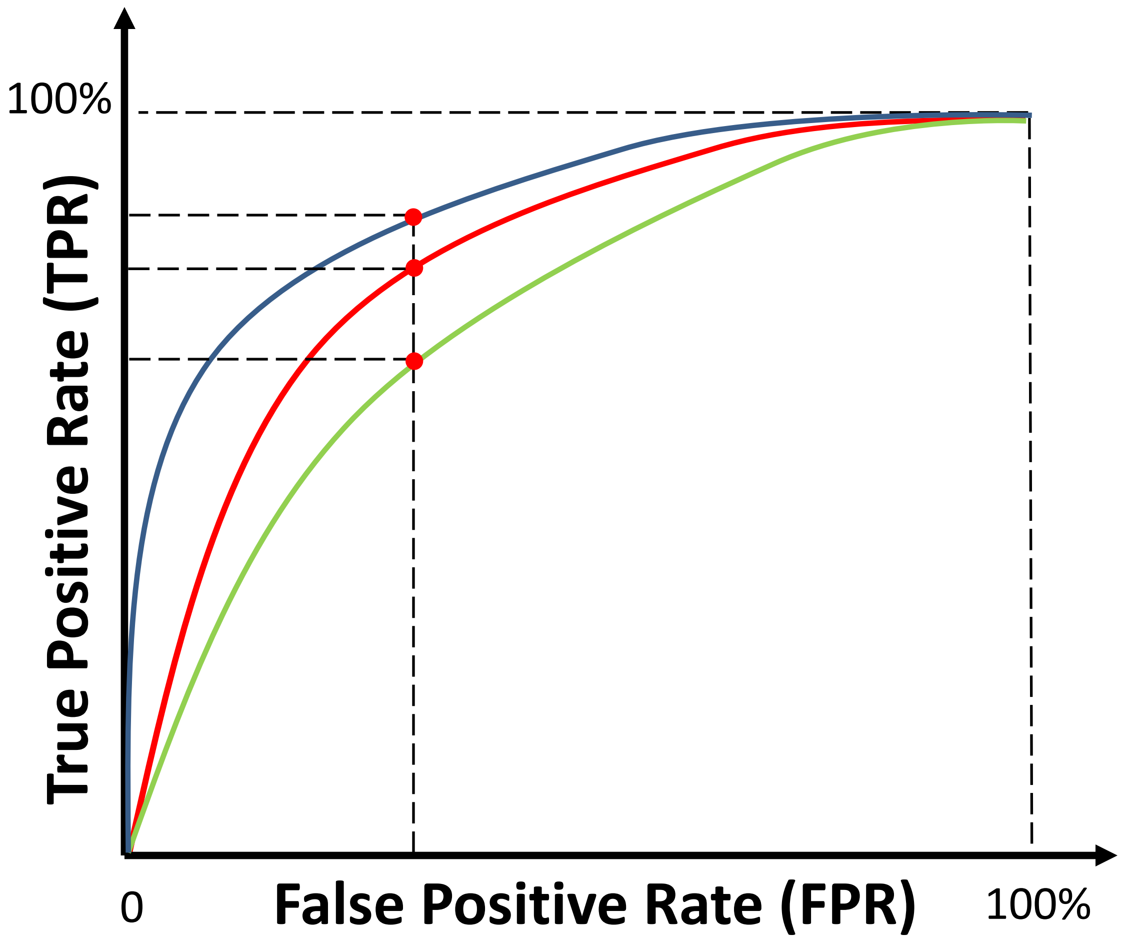


ROC Curve





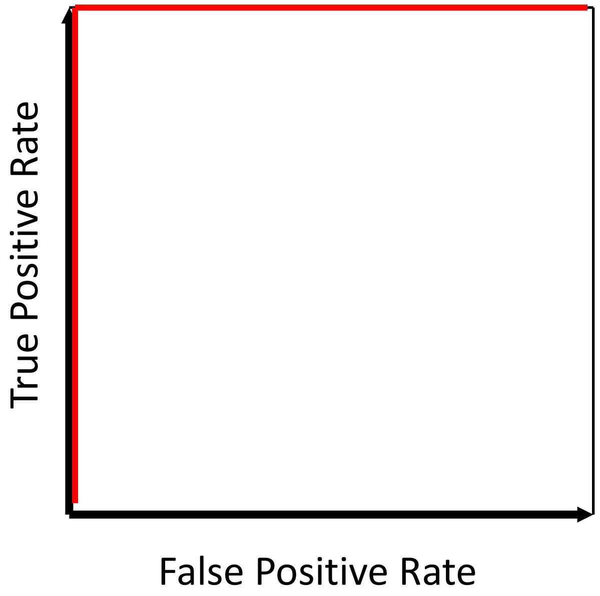
Question: Which one is better?



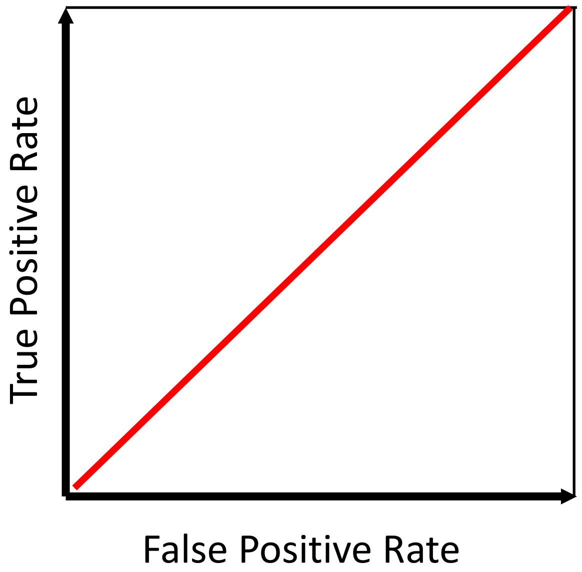
Question: Which one is better?

Answer: The blue one! Because it provides higher TPR for a fixed FPR.

Special Cases



Best Case: 100% Accuracy!
(FPR = 0, TPR = 100%)

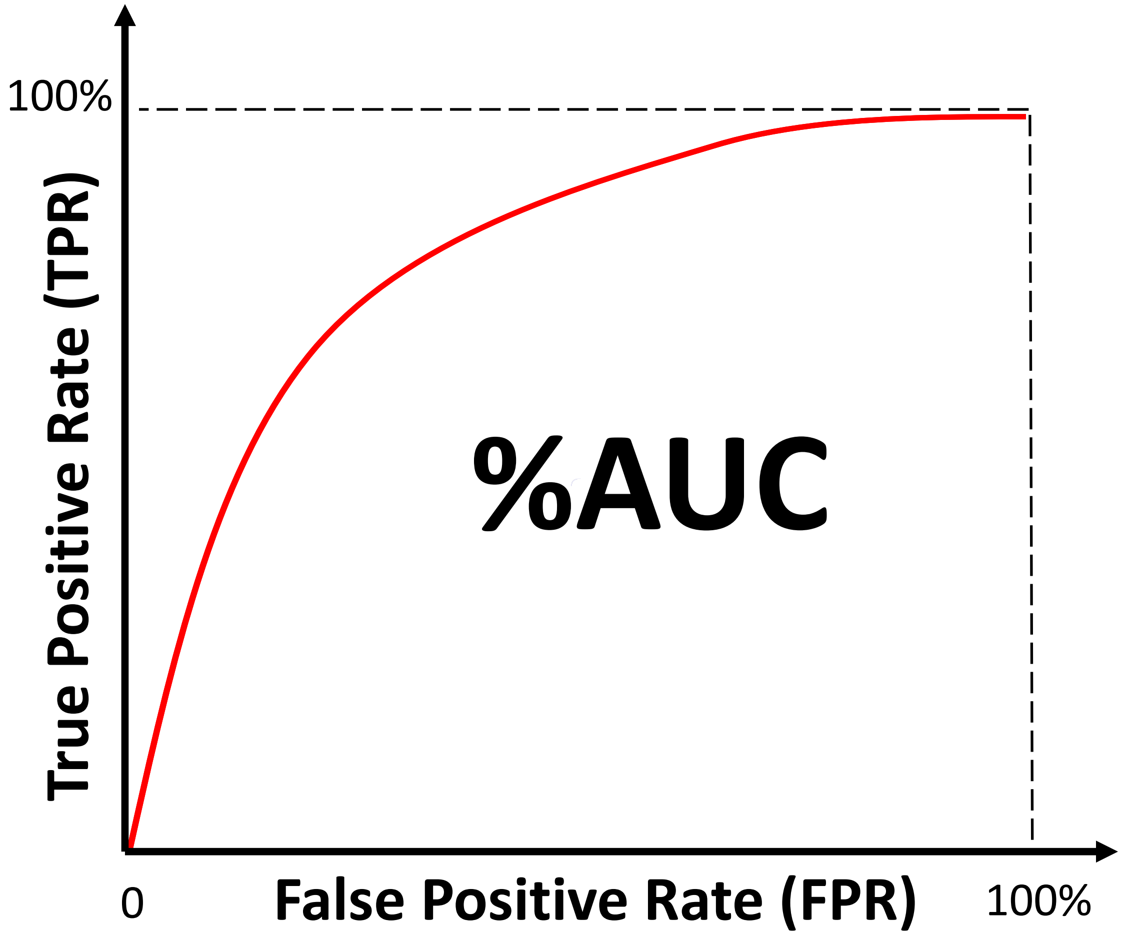


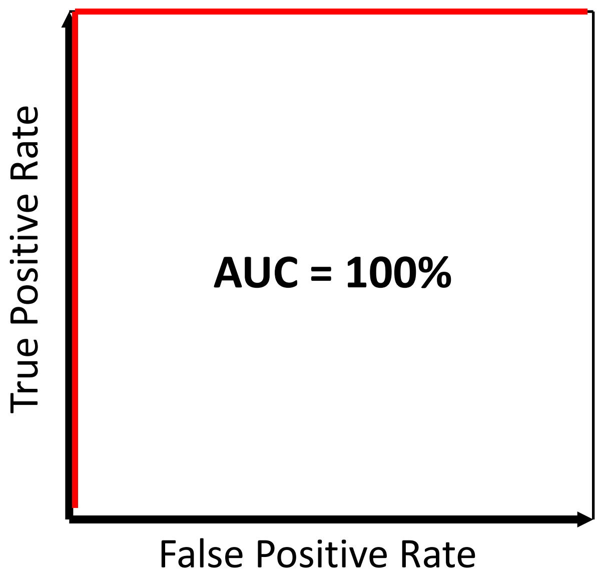
Worst Case: Random Guess!
(FPR = TPR)

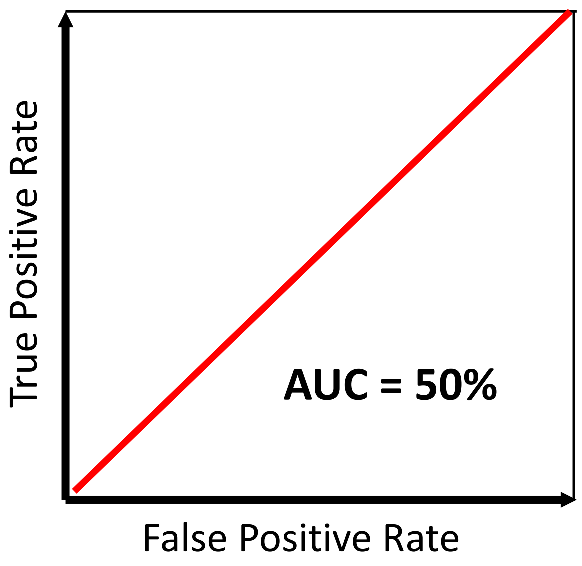
AUC (Area Under Curve)

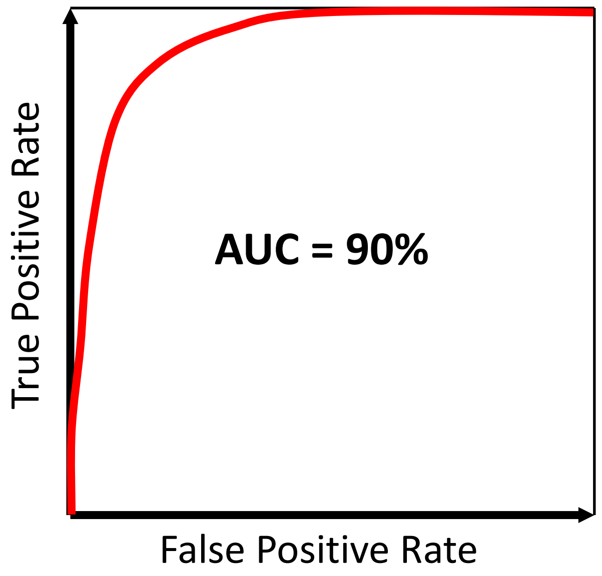
- As we saw, the more closer to the upper left corner, the better!
- So, the **Area Under Curve (AUC)** for ROC can be a good metric to represent the overall performance of a classifier!

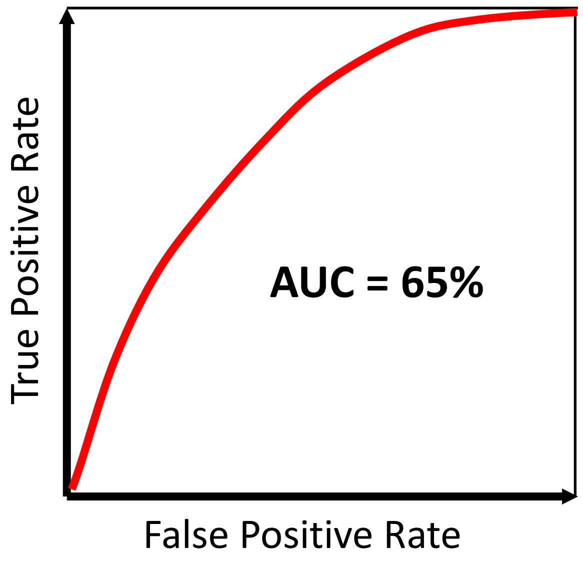
AUC for ROC Curve




AUC = 100%
True Positive Rate
False Positive Rate


AUC = 50%
True Positive Rate
False Positive Rate


AUC = 90%
True Positive Rate
False Positive Rate


AUC = 65%
True Positive Rate
False Positive Rate

Thank You!

Questions?