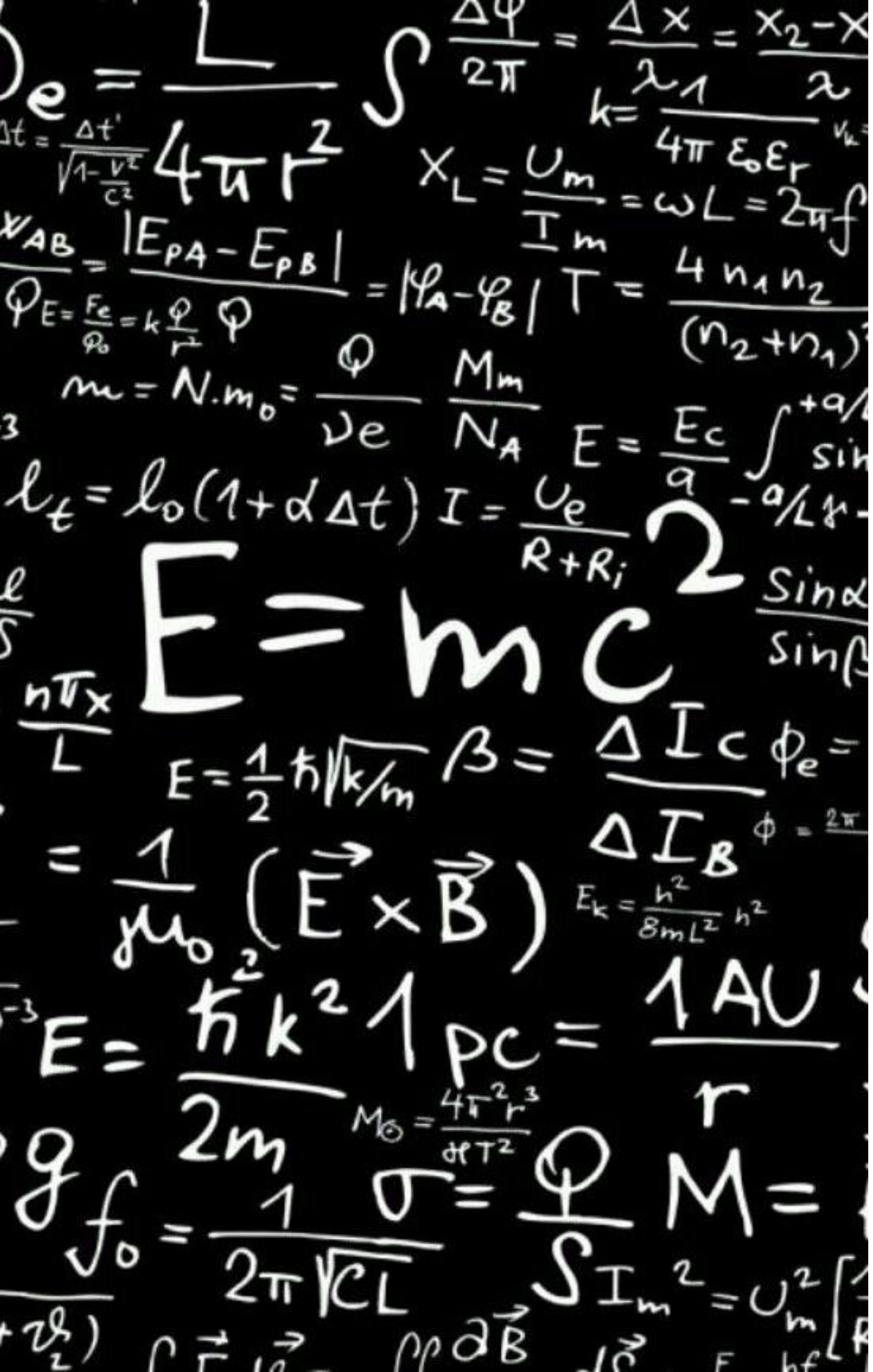
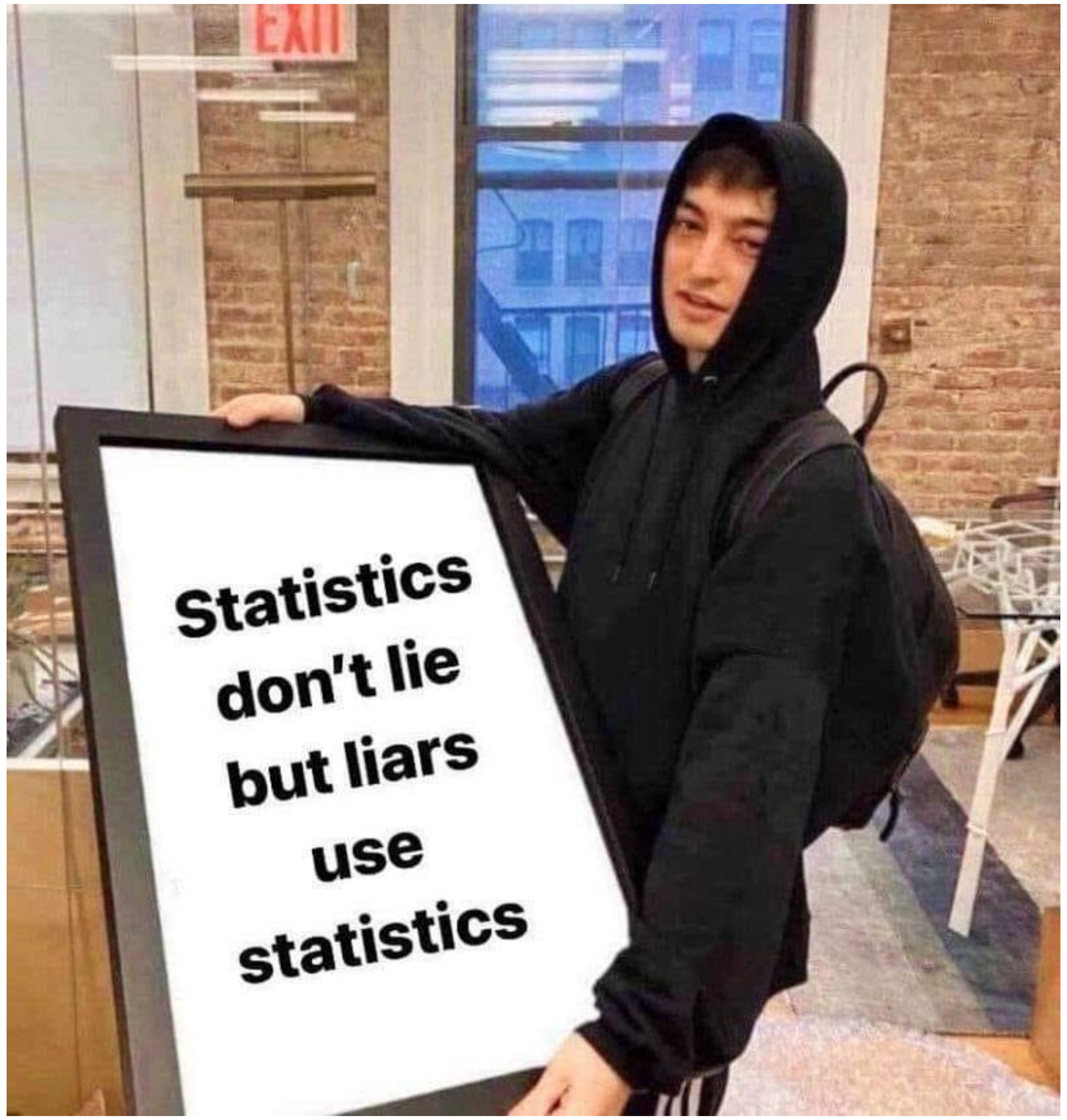


RESULT METRICS and the ANALYSIS

A Quote



https://www.facebook.com/iamaprogrammerbro/photos/a.699904080404992/1341502592911801/

A Quote

We have achieved 99% accuracy on our COVID-19 patient screening.

A Quote

We have achieved 99% accuracy on our COVID-19 patient screening app.

What is wrong with this quote?

A Quote

We have achieved 99% accuracy on our COVID-19 patient screening app.

What is wrong with this quote?

- Sample size
- Input Distribution
- Result Distribution

Metrics

Different problems require different metrics

- ➤ Regression problem
- ➤ Classification problem
- ➤ Object Detection

Metrics

Different problems require different metrics

- Regression problem
  - Sum of squared error (SSE)
  - Mean squared error (MSE)
  - Root mean squared error (RMSE)
- Classification problem
- Object Detection

Metrics

Different problems require different metrics

- Regression problem
- Classification problem
  - Confusion Matrix
  - Accuracy, Precision, Recall, F1-Score
  - ROC
- Object Detection

Metrics

Different problems require different metrics

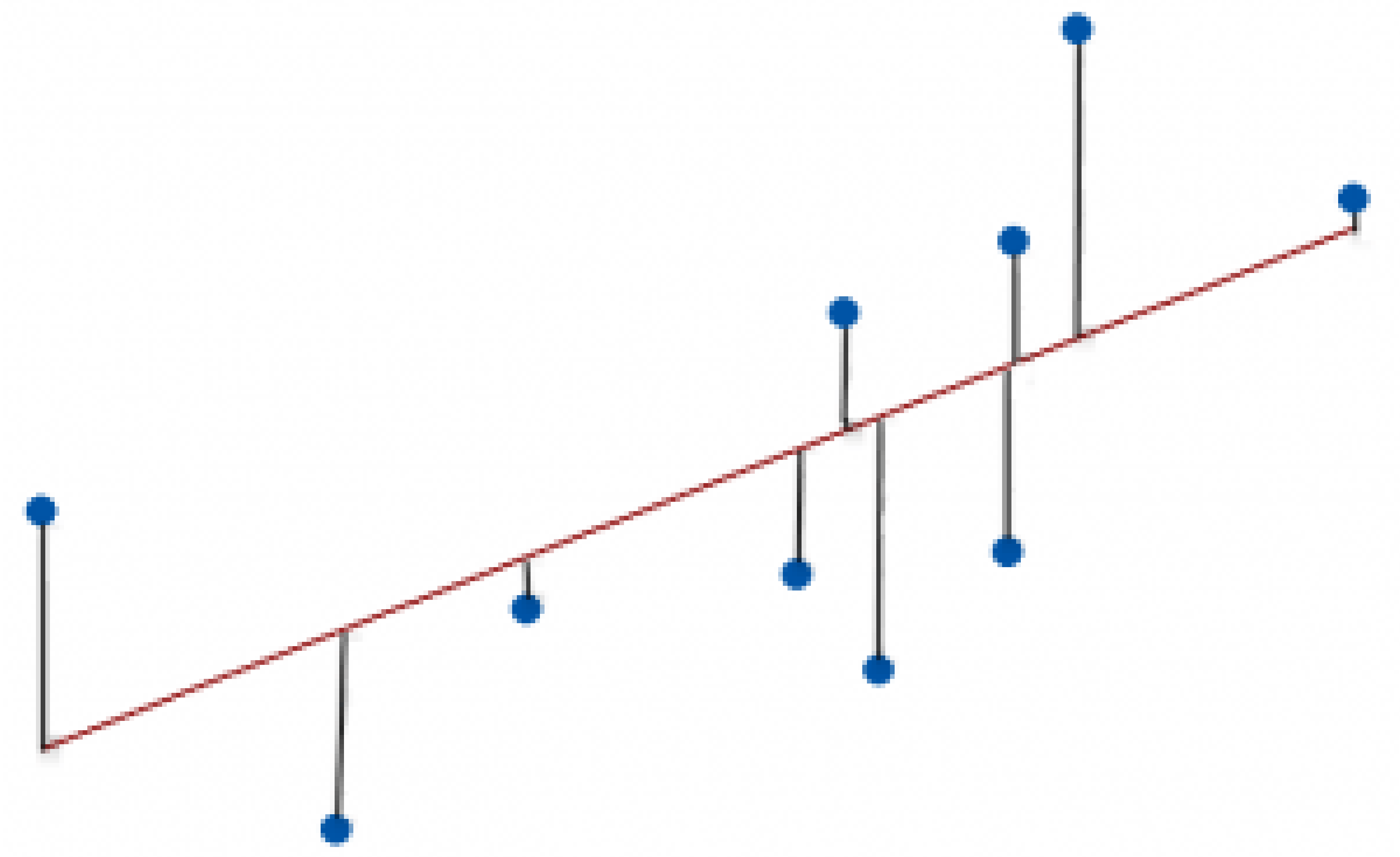
- Regression problem
- Classification problem
- Object Detection (we will cover this topic in Deep Learning course)
  - Intersect over Union (IOU)
  - Mean Average Precision (mAP)

Regression Metrics

Different problems require different metrics

- Regression problem
  - Sum of squared error (SSE)
  - Mean squared error (MSE)
  - Root mean squared error (RMSE)
- Classification problem
- Object Detection

Regression Metrics - SSE



Sum of squared error

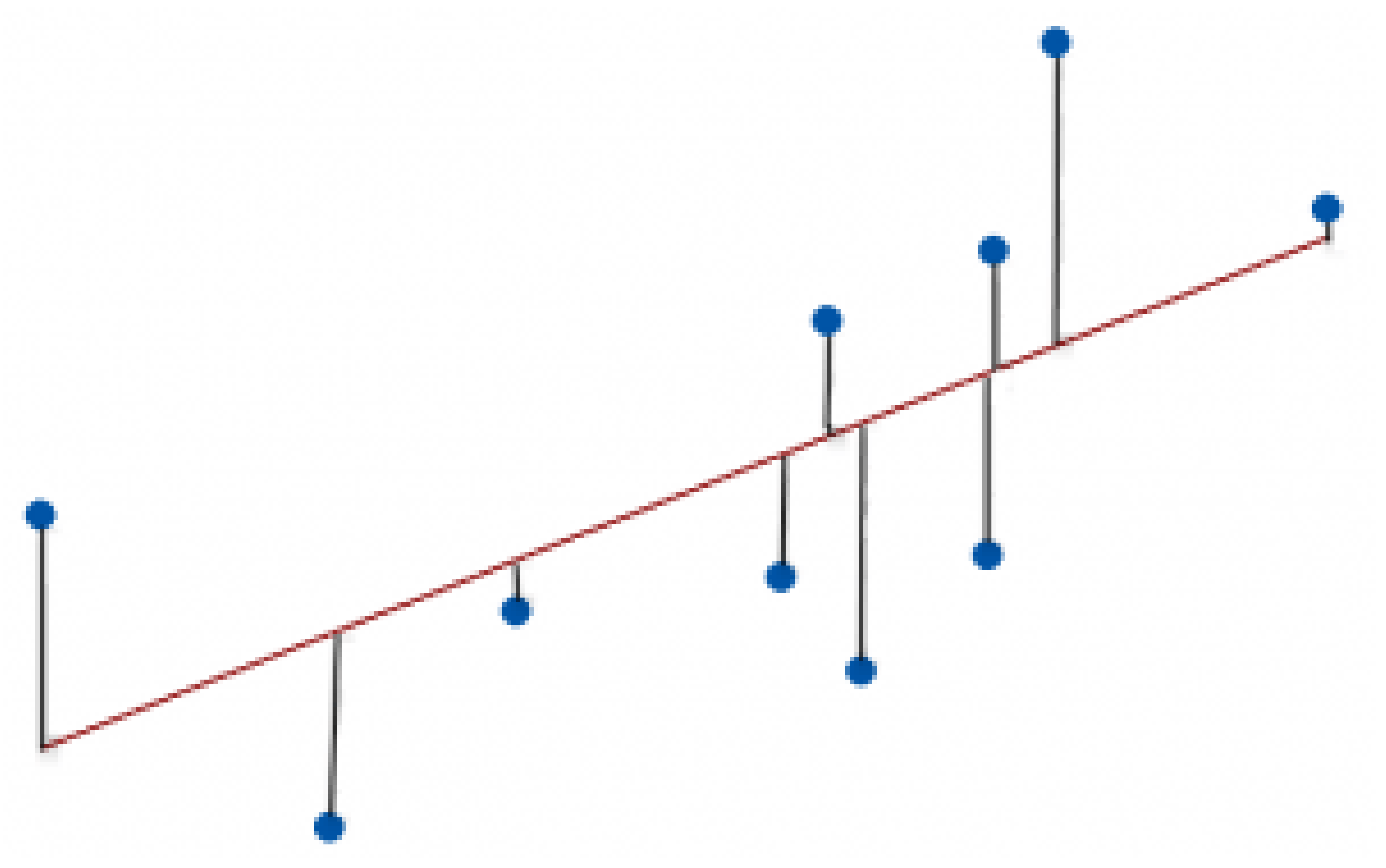
$$SSE = \sum (y - \hat{y})^2$$

y = real output

$\hat{y} = h(\vec{x})$ = predicted output

https://i0.wp.com/statisticsbyjim.com/wp-content/uploads/2017/03/residuals-300x186.png?resize=300%2C186&is-pending-load=1

Regression Metrics - MSE



Mean squared error

$$MSE = \frac{1}{n}\sum(y - \hat{y})^2$$

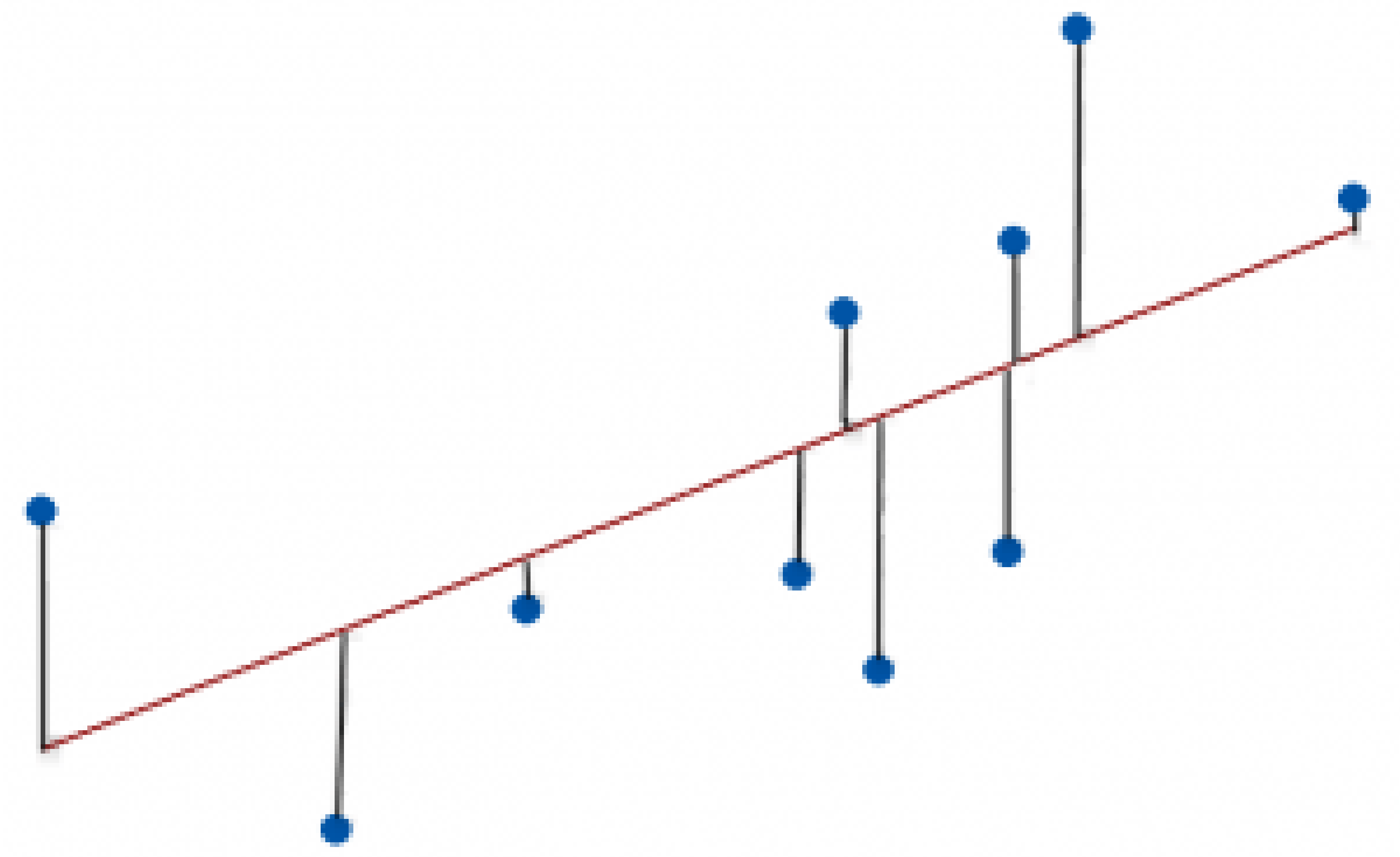
y = real output

$\hat{y} = h(\vec{x})$ = predicted output

n = number of instances

https://i0.wp.com/statisticsbyjim.com/wp-content/uploads/2017/03/residuals-300x186.png?resize=300%2C186&is-pending-load=1

Regression Metrics - RMSE



Root mean squared error

$$RMSE = \sqrt{\frac{1}{n}\sum(y - \hat{y})^2}$$

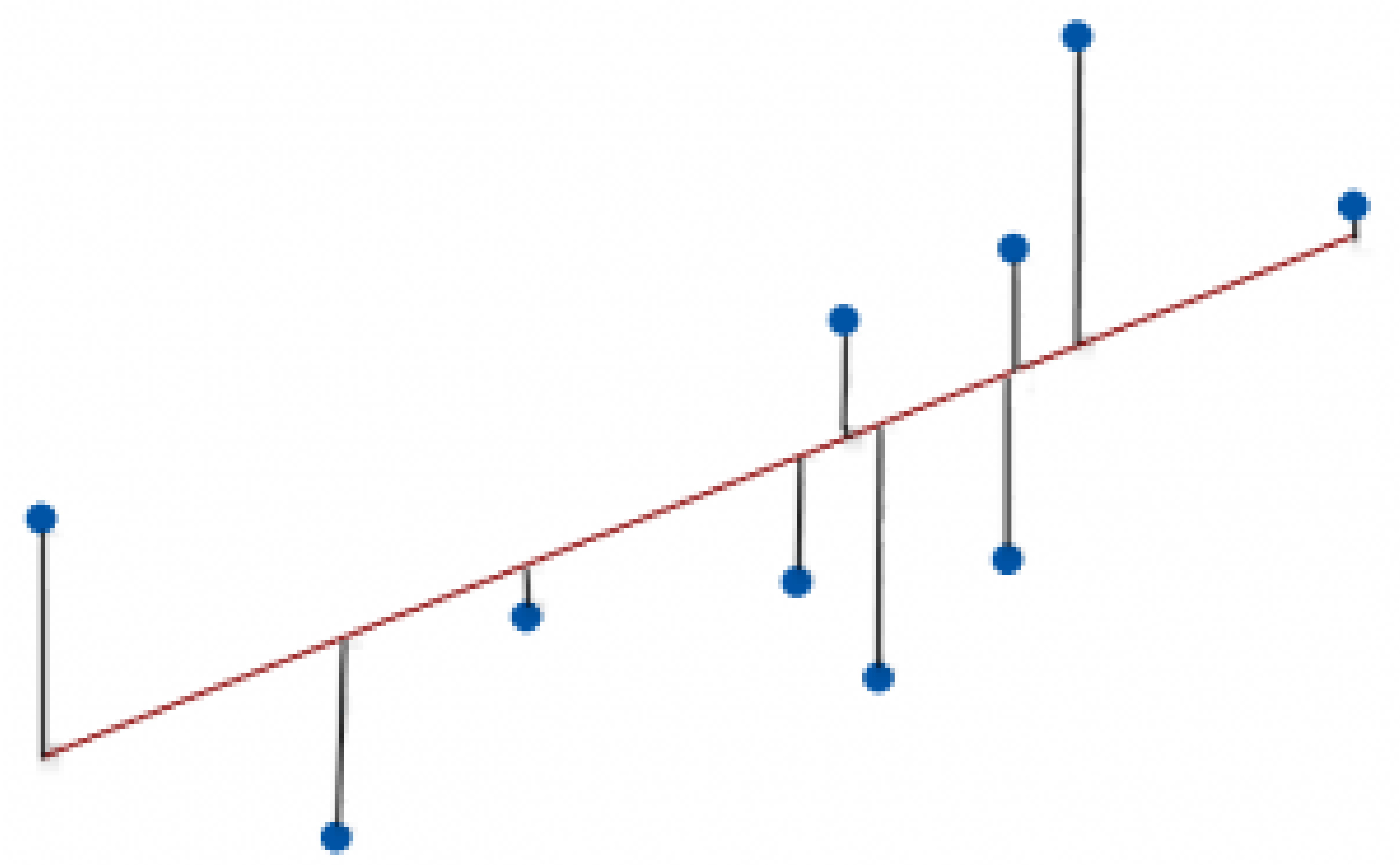
y = real output

$\hat{y} = h(\vec{x})$ = predicted output

n = number of instances

https://i0.wp.com/statisticsbyjim.com/wp-content/uploads/2017/03/residuals-300x186.png?resize=300%2C186&is-pending-load=1

Regression Metrics



What do they tell us?

SSE, MSE, RMSE of one model?

Comparing SSE, MSE, RMSE of multiple model?

https://i0.wp.com/statisticsbyjim.com/wp-content/uploads/2017/03/residuals-300x186.png?resize=300%2C186&is-pending-load=1

Classification Metrics

Confusion Matrix

A table that explains the result of predictions compared to the actual result.

Predict = What the algorithm predicts

Actual = The actual value

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Classification Metrics - Confusion Matrix

Confusion Matrix

A table that explains the result of predictions compared to the actual result.

Positive Actual Value = number of positive instances

Negative Actual Value = number of negative instance

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Classification Metrics - Confusion Matrix

Confusion Metrics

A table that explains the result of predictions compared to the actual result.

Positive Prediction = number of instances the algorithm predicts as positive

Negative Prediction = number of instances the algorithm predicts as negative

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Classification Metrics - Confusion Matrix

True Positive

➤ Positive value that is predicted as positive

True Negative

➤ Negative value that is predicted as negative

False Positive (Type I error)

➤ Negative value that is predicted as positive

False Negative (Type II error)

➤ Positive value that is predicted as negative

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Classification Metrics - Confusion Matrix

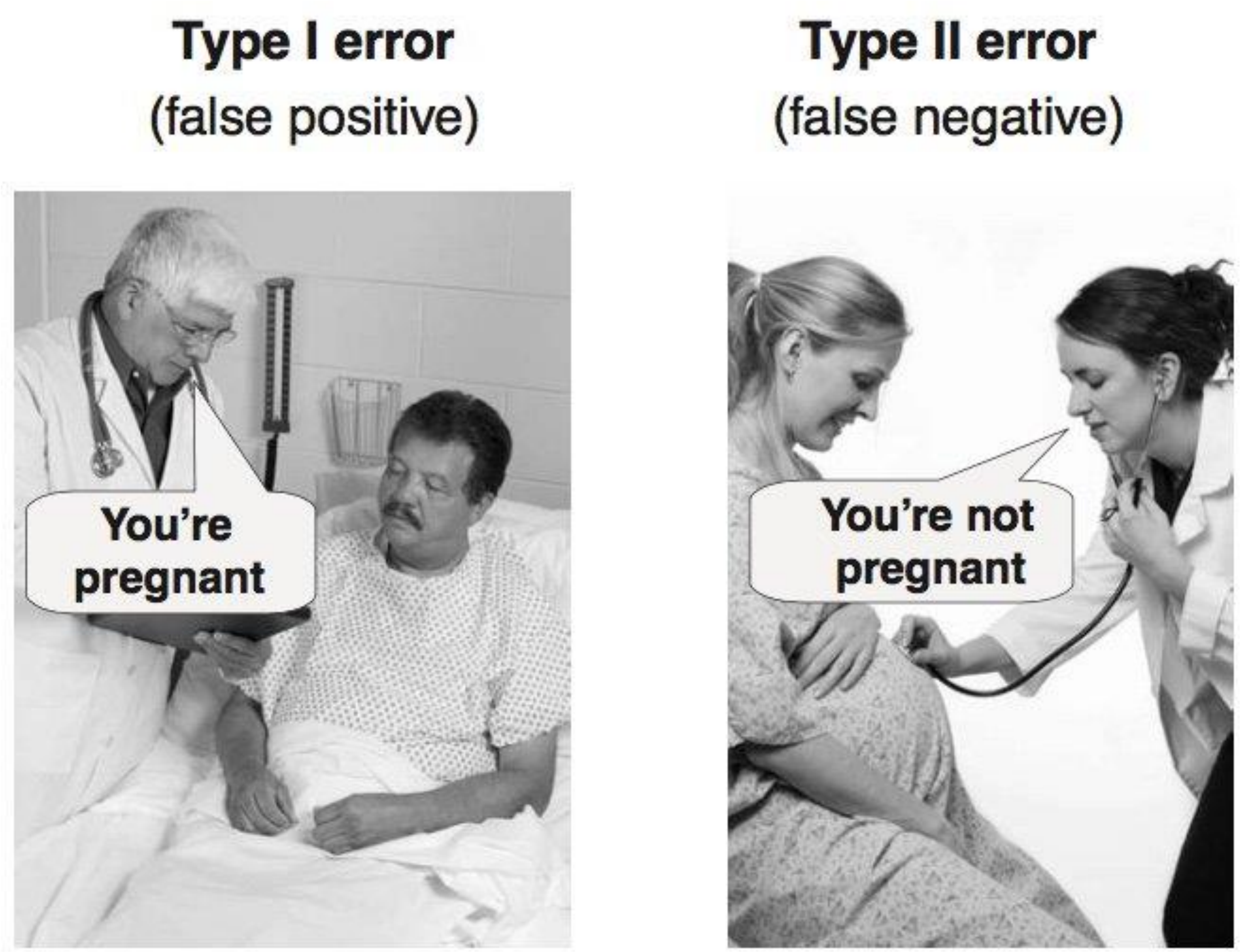


Image from The Essential Guide to Effect Sizes by Paul D Ellis.

Classification Metrics – Acc, Pre, Rec

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

What percentage can the algorithm predict correctly?

$$Precision = \frac{TP}{TP + FP}$$

Out of all predicted positives, what percentage can the algorithm predict correctly?

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

Out of all actual positives, what percentage can the algorithm predict correctly?

		Actual	
		Positive	**Negative**
Predicted	**Positive**	**True Positive**	**False Positive**
	Negative	**False Negative**	**True Negative**

Classification Metrics

Let’s try to understand this:

Company S has developed a test kit for screening virus C. Upon testing, its confusion matrix looks like this.

		Actual	
		Positive	Negative
Predicted	Positive	60	50
	Negative	40	850

a) How many tests are there?
b) How many positives?
c) How many negatives?
d) Accuracy?
e) Precision?
f) Recall?
g) Should we use it?

Classification Metrics

Let's try to understand this:

Company S has developed a test kit for screening virus C. Upon testing, its confusion matrix looks like this.

		Actual	
		Positive	Negative
Predicted	Positive	60	50
	Negative	40	850

In medical field, scientists need more tools to analyze.

$$Sensitivity = True\ Positive\ Rate = \ Recall = \frac{TP}{TP + FN}$$

$$Specificity = True\ Negative\ Rate = \frac{TN}{TN + FP}$$

Classification Metrics – F1 Score

$$F_1 = \frac{2 * precision * recall}{precision + recall}$$

F_1 score measures a test's accuracy using the combination of precision and recall.

It ranges between 0 and 1.

- ➤ 0 if either precision or recall is zero
- ➤ 1 if the prediction has perfect precision and recall

		Actual	
		Positive	**Negative**
Predicted	**Positive**	**True Positive**	**False Positive**
	Negative	**False Negative**	**True Negative**

Classification Metrics – ROC Curve

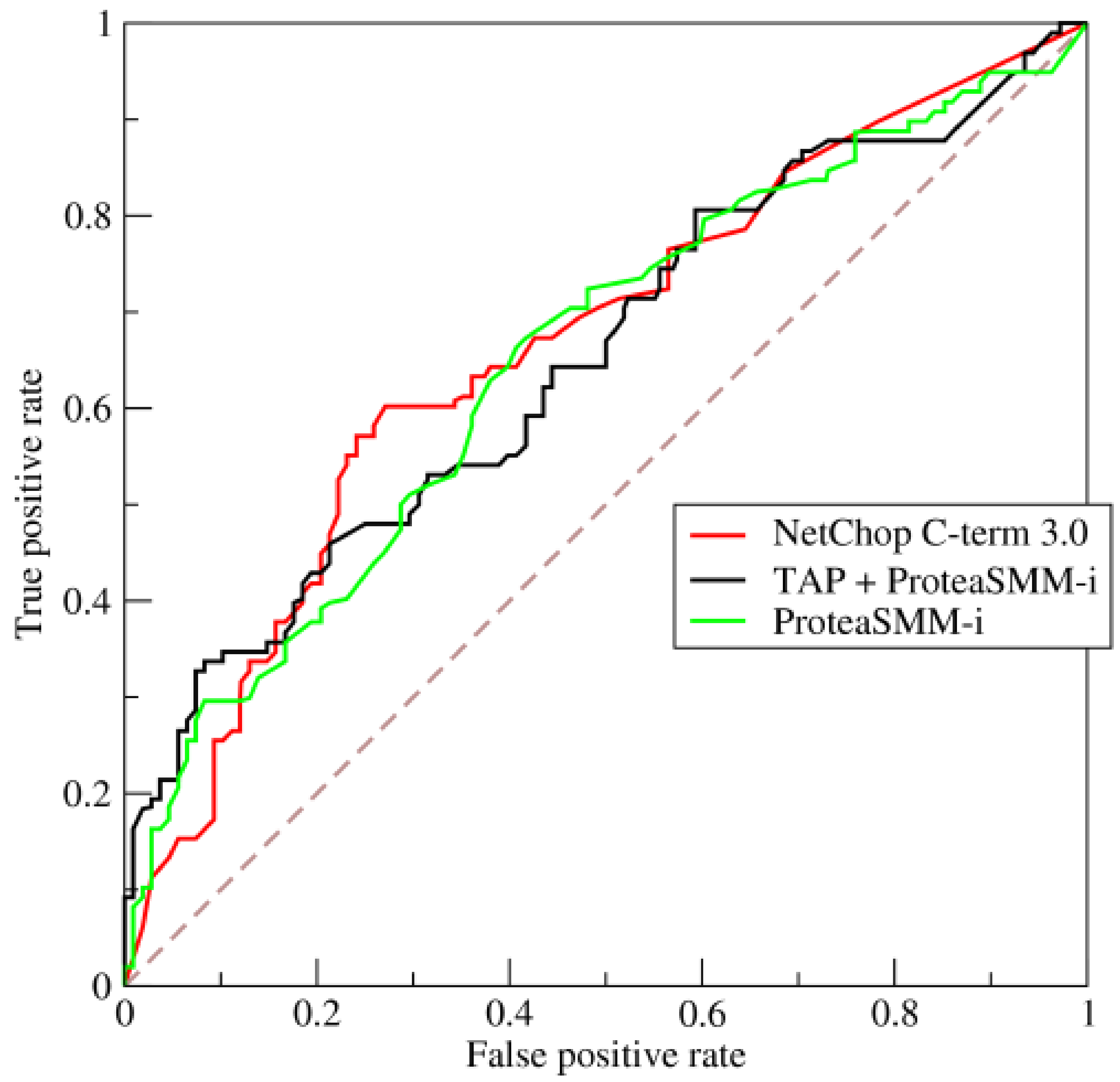
ROC Curve (Receiver Operating Characteristics Curve)

➤ A curve that shows relationships between TPR and FPR when discrimination threshold is varied.

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

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

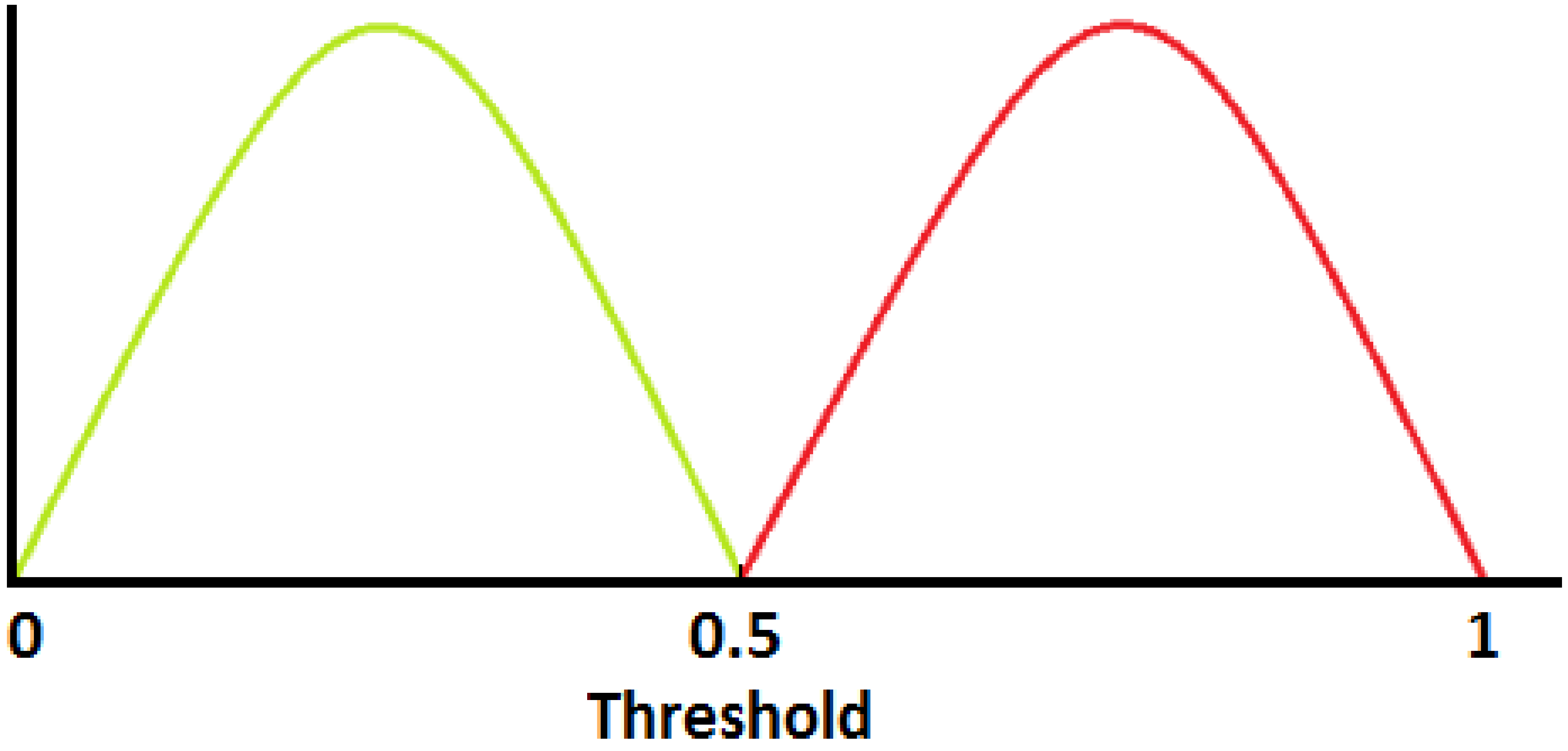
		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative



https://en.wikipedia.org/wiki/Receiver_operating_characteristic#/media/File:Roccurves.png

Classification Metrics – ROC Curve

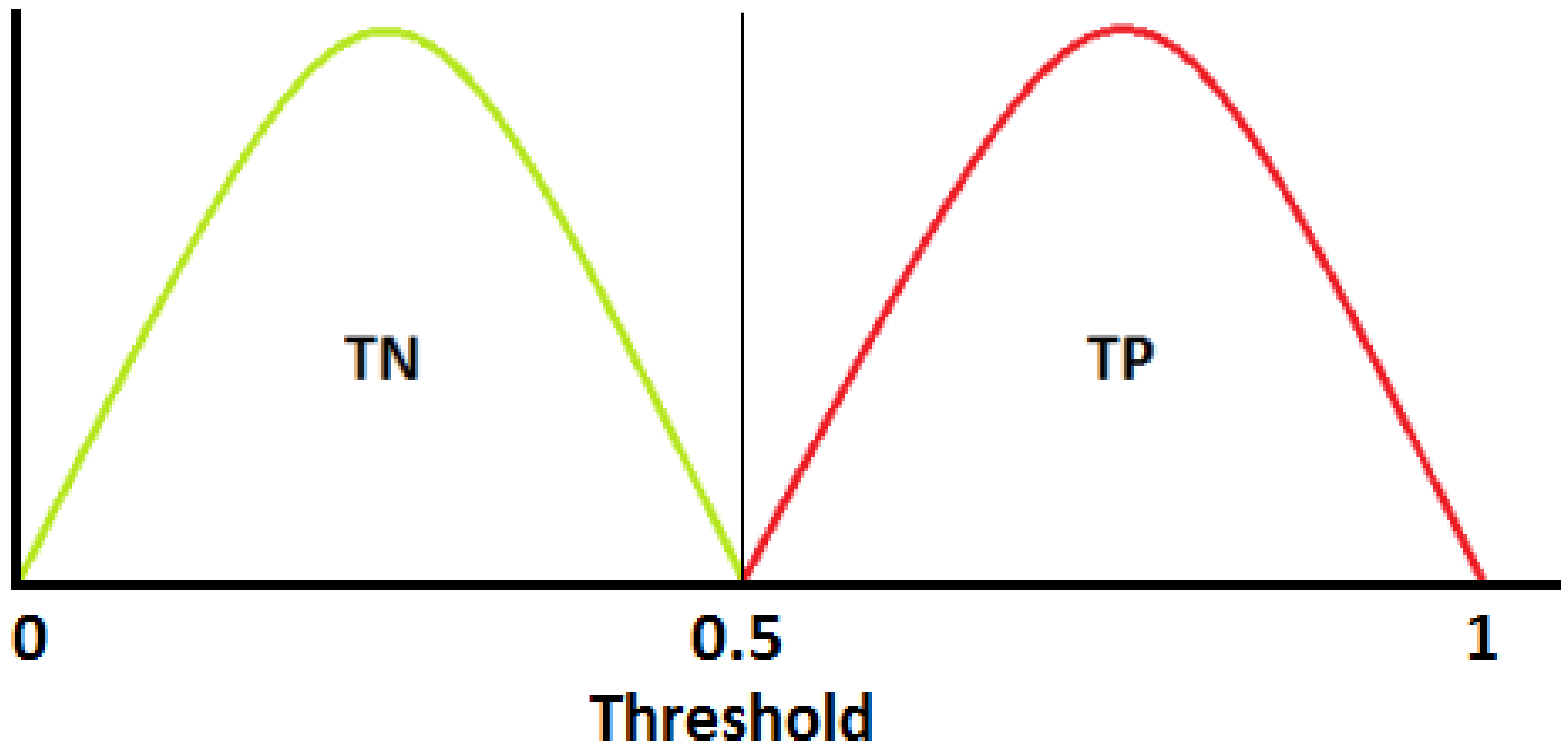
Let’s look at how thresholds affect results.



		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Classification Metrics – ROC Curve

If we select threshold = 0.5, the result will look like this.



		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

$$TPR = \frac{TP}{TP + FN} \qquad FPR = \frac{FP}{FP + TN}$$

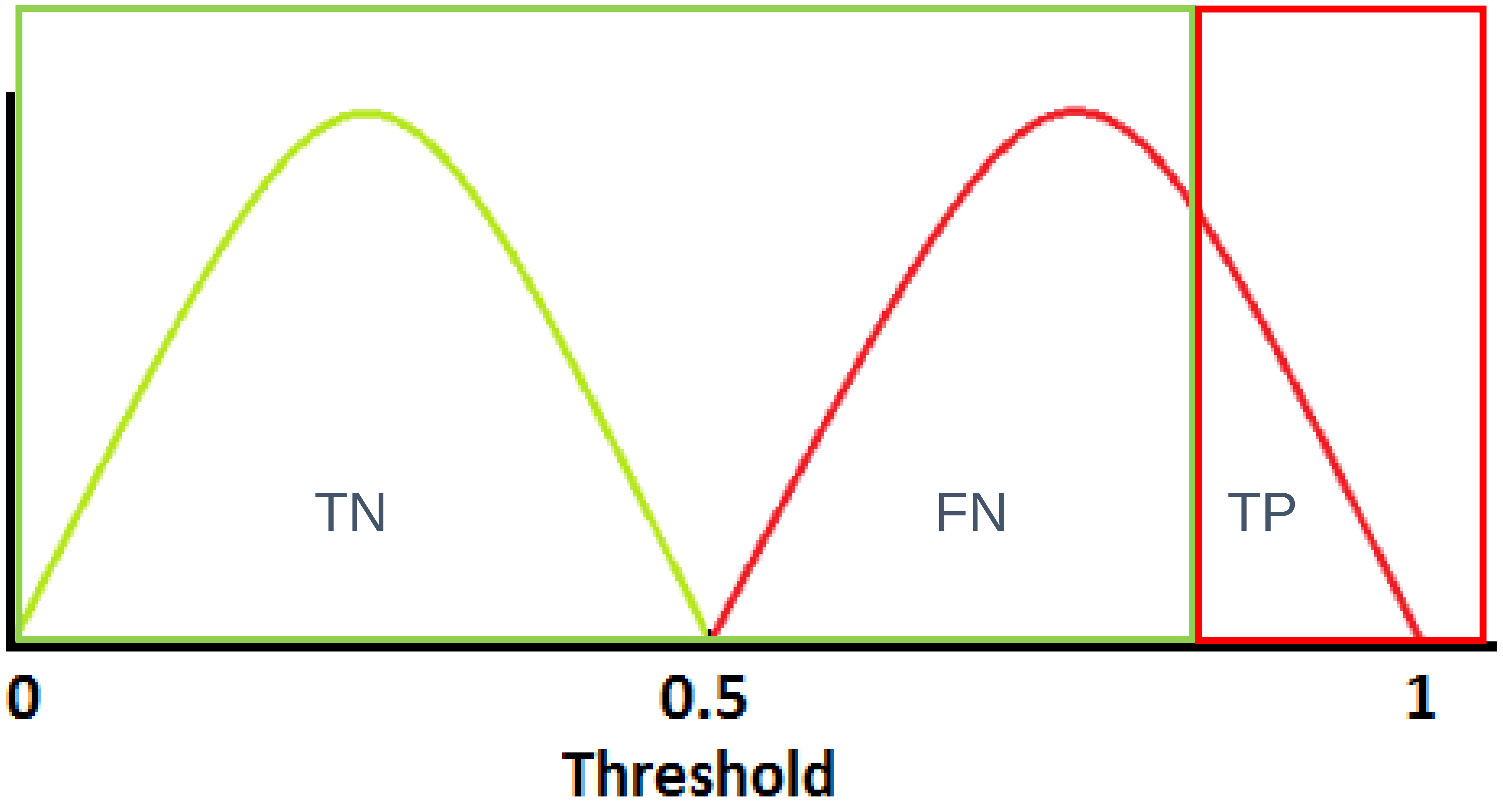
TPR = 1

FPR = 0

https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5

Classification Metrics – ROC Curve

If we select threshold = 0.8, the result will look like this.



		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

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

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

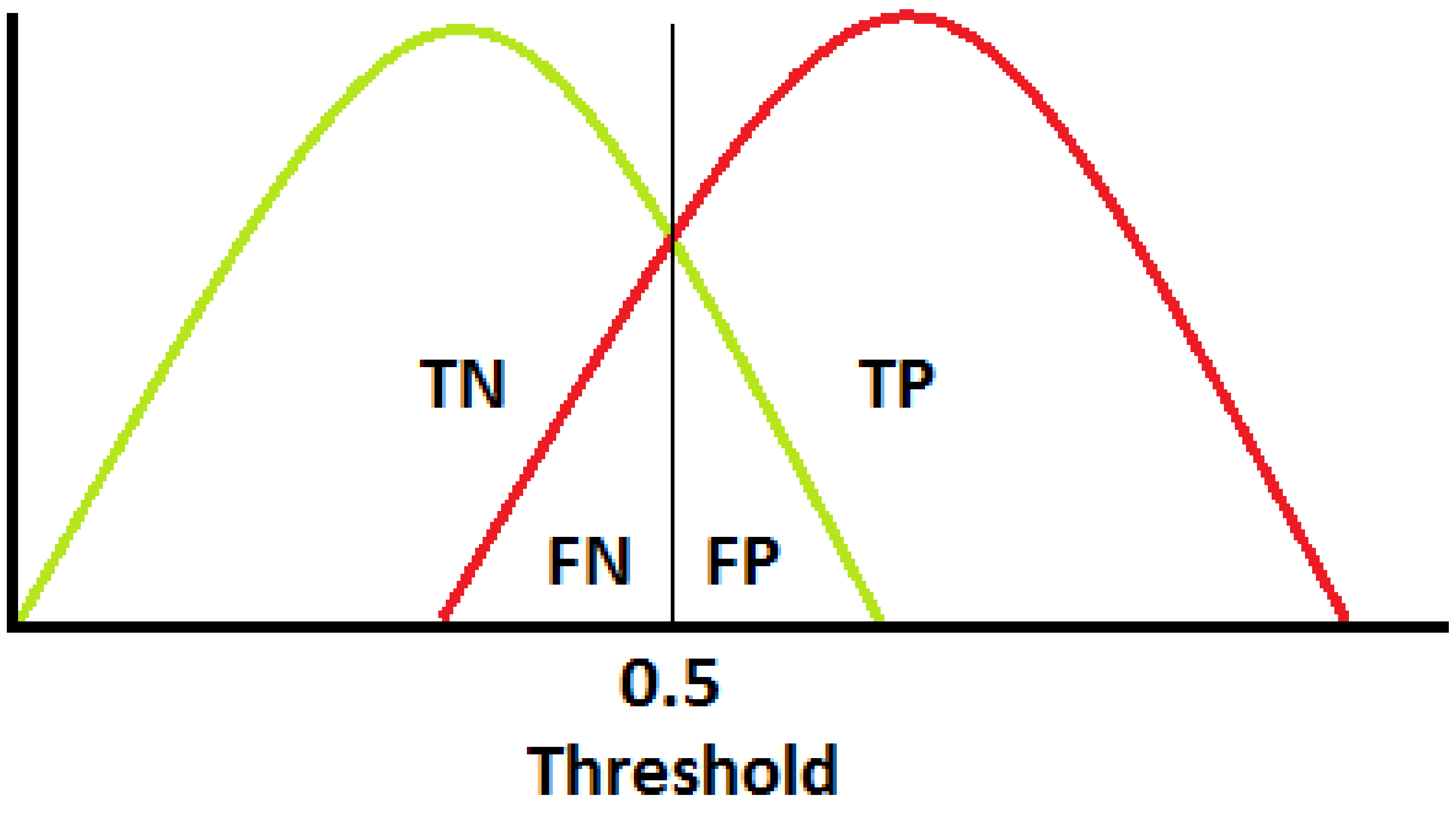
TPR < 1

FPR = 0

https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5

Classification Metrics – ROC Curve

What if the data distributions overlap?



		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

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

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

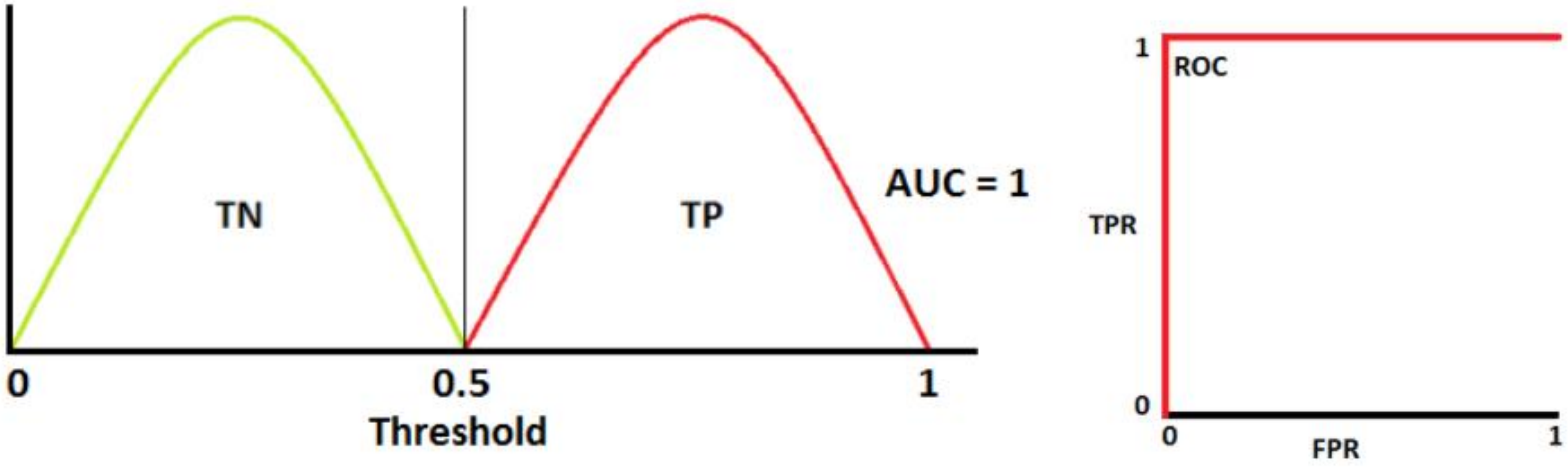
TPR < 1

0 < FPR < 1

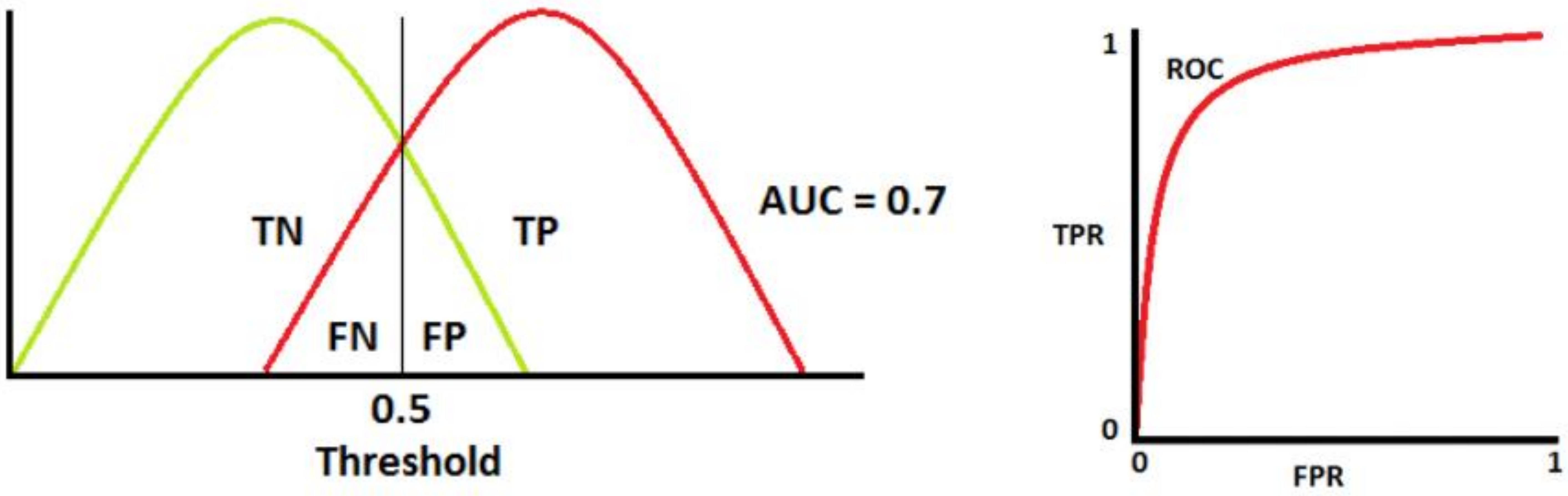
https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5

Classification Metrics – ROC Curve

Perfect ROC



ROC that we will mostly get

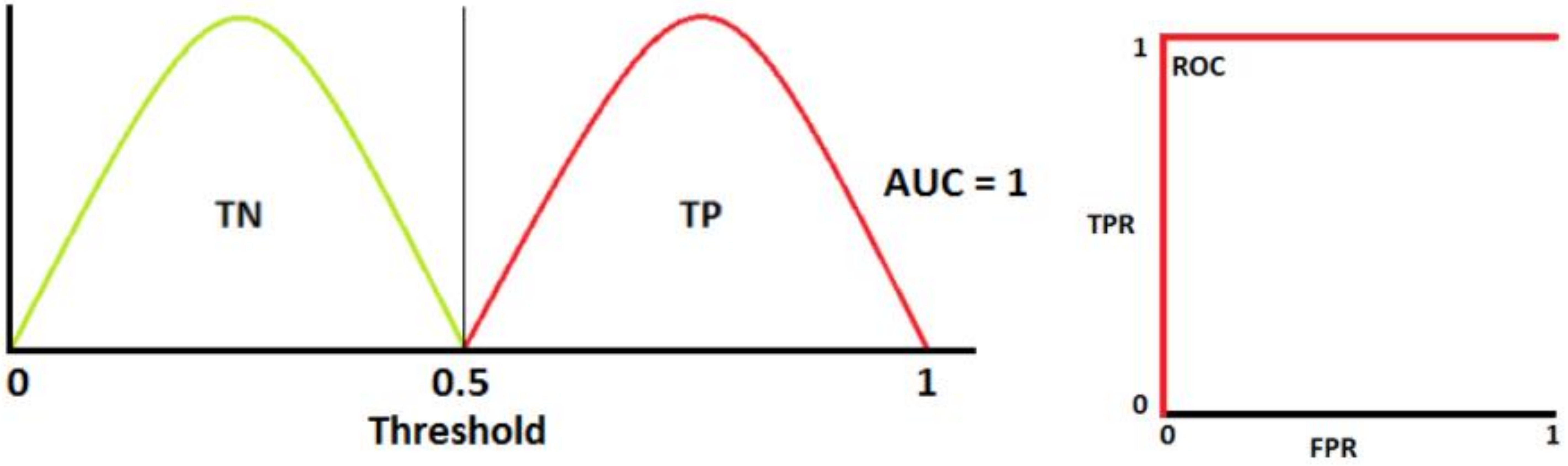


https://en.wikipedia.org/wiki/Receiver_operating_characteristic#/media/File:Roccurves.png

Classification Metrics – ROC Curve

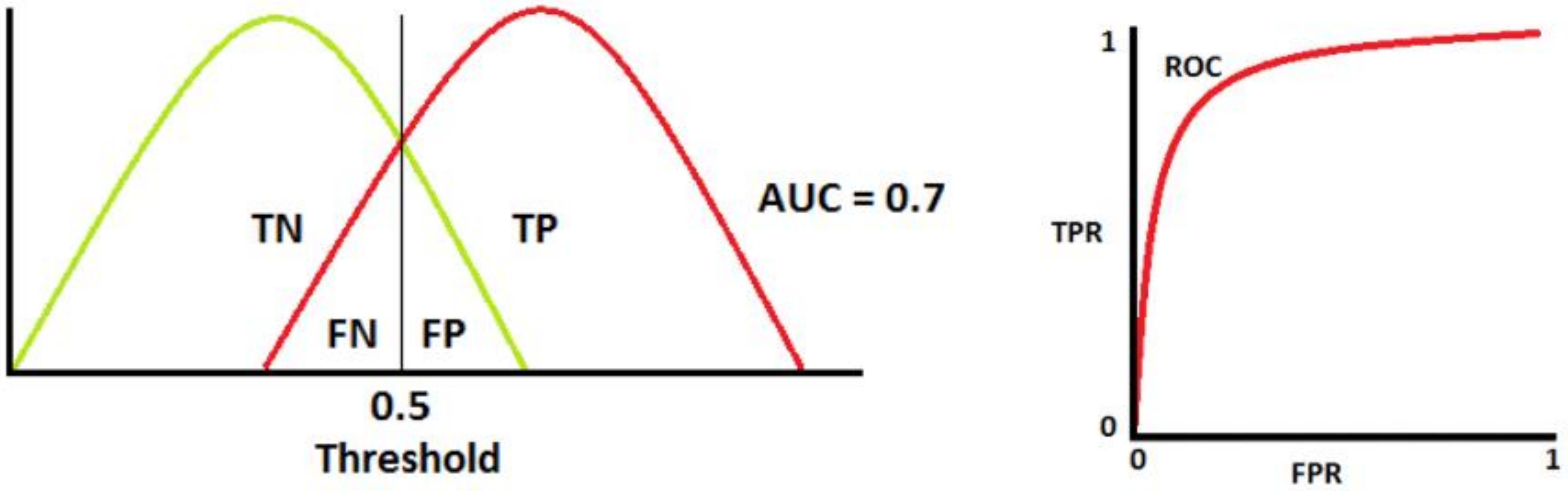
Perfect ROC

➤ Area Under Curve (AUC) is 1



ROC that we will mostly get

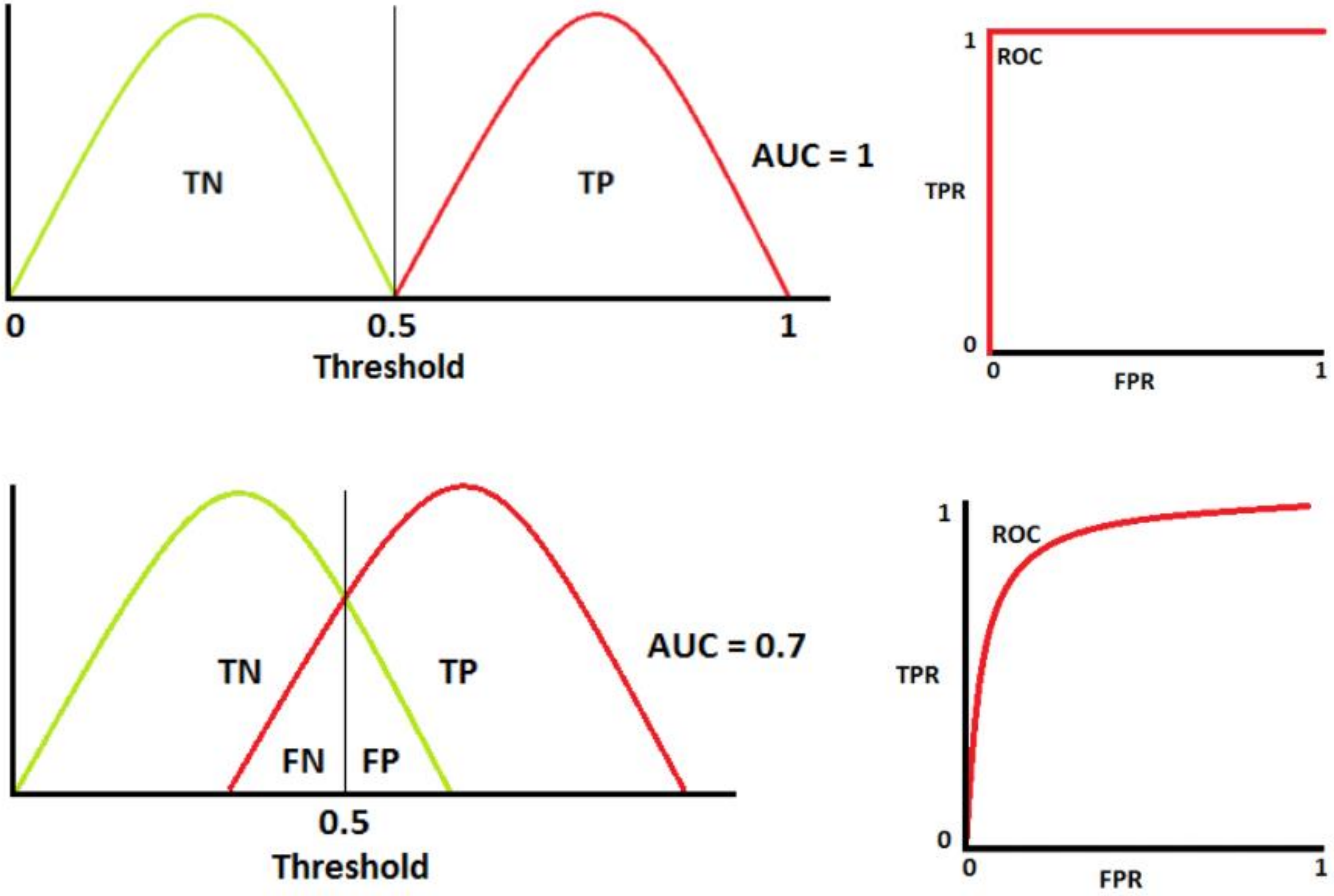
➤ Area Under Curve (AUC) is < 1



https://en.wikipedia.org/wiki/Receiver_operating_characteristic#/media/File:Roccurves.png

Classification Metrics – AUC

Given 2 models, model with higher AUC performs better.



https://en.wikipedia.org/wiki/Receiver_operating_characteristic#/media/File:Roccurves.png

Back to the Quote

We have achieved 99% accuracy on our COVID-19 patient screening app.

What is wrong with this quote?

As an ML engineer, how should you train a model to NOT output the quote above?