IRON HACK

EVALUATION METRICS

PERFORMANCE EVALUATION METRICS - ACCURACY

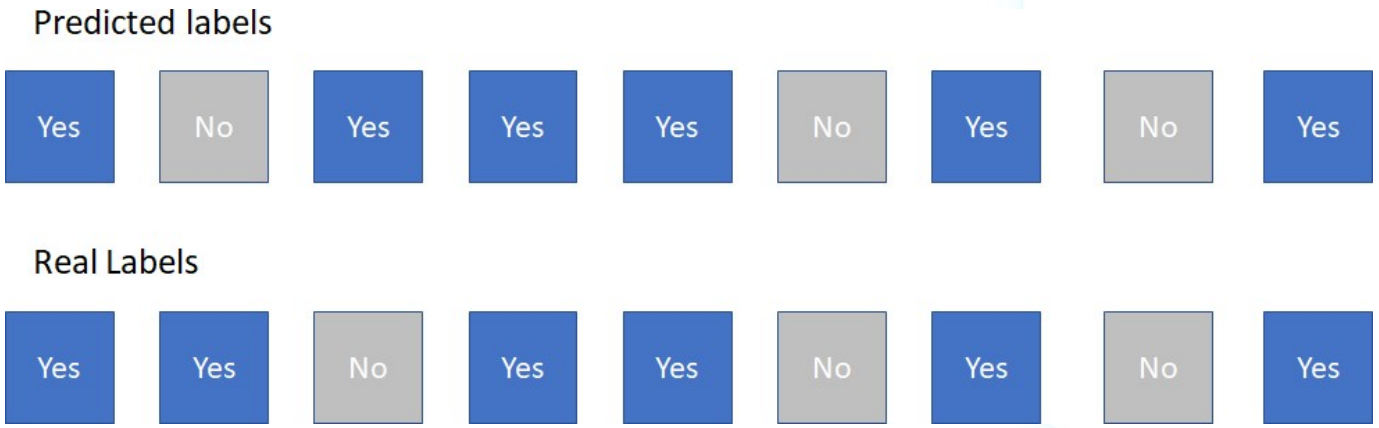
So far we have just been evaluating the performance of our algorithms based on accuracy

$$\frac{\#\text{correct Prediction}}{\text{Total \# of Predictions}}$$

PERFORMANCE EVALUATION METRICS - ACCURACY

So far we have just been evaluating the performance of our algorithms based on accuracy

$$\frac{\#\text{correct Prediction}}{\text{Total \# of Predictions}}$$



PERFORMANCE EVALUATION METRICS - ACCURACY

So far we have just been evaluating the performance of our algorithms based on accuracy

$$\frac{\#\text{correct Prediction}}{\text{Total \# of Predictions}}$$

PERFORMANCE EVALUATION METRICS - ACCURACY

So far we have just been evaluating the performance of our algorithms based on accuracy

#correct Prediction

Total # of Predictions

$$Accuracy = \frac{\#\ correct\ predictions}{\#\ total\ predictions} = \frac{6}{9} \sim 66\%$$

Predicted labels

Yes	No	Yes	Yes	Yes	No	Yes	No	Yes

Real Labels

Yes	Yes	No	Yes	Yes	No	Yes	No	No

PERFORMANCE EVALUATION METRICS - ACCURACY

How can accuracy be deceptive?

- **Imbalanced Datasets**

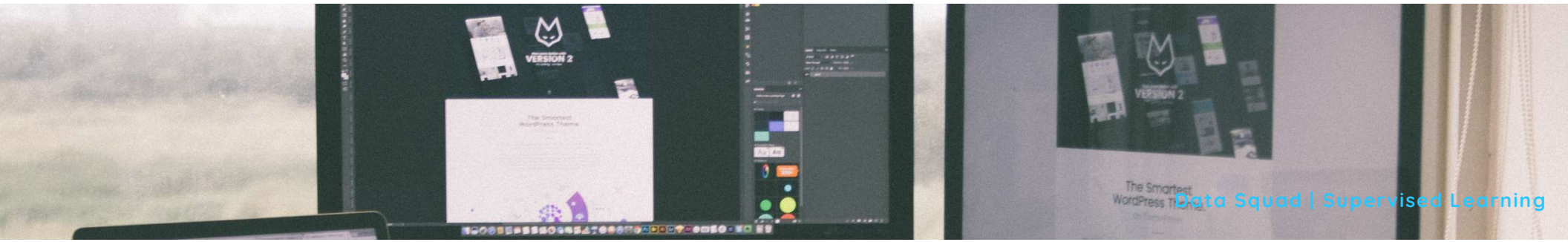
- **Importance of incorrect prediction of a certain class**

$$\frac{\#\text{correct Prediction}}{\text{Total \# of Predictions}}$$

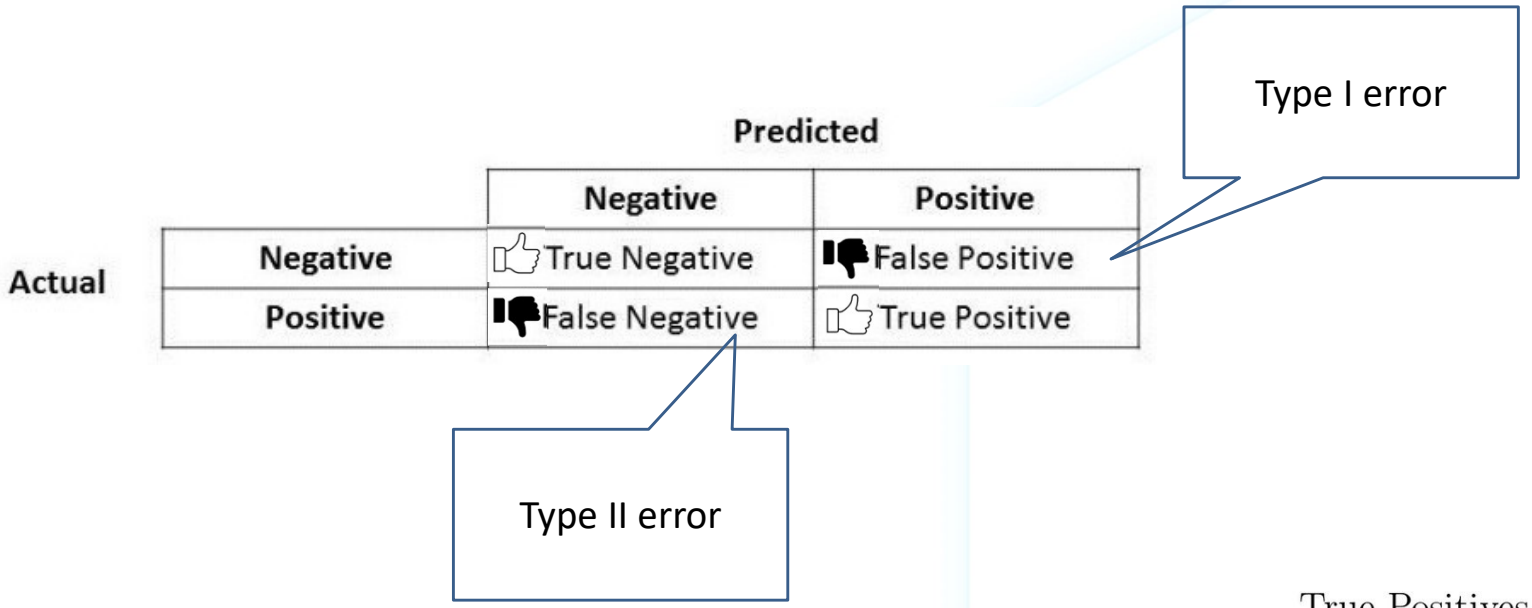
PERFORMANCE EVALUATION METRICS – CONFUSION MATRIX

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

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{All predictions}}$$



PERFORMANCE EVALUATION METRICS – CONFUSION MATRIX



$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{All predictions}}$$

PERFORMANCE EVALUATION METRICS – CONFUSION MATRIX

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

Type I error

Type II error

Type I Error

Type II Error

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{All predictions}}$$

PERFORMANCE EVALUATION METRICS – CONFUSION MATRIX

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

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{All predictions}}$$

PERFORMANCE EVALUATION METRICS - PRECISION & RECALL

Let's describe these two concepts through words & intuition

Precision of Class A:

From all the datapoints I predicted to be "A", how many were A?

Example: From a universe of 1000 people, we predicted 10 had coronavirus. In fact only 8 out of these 10 were actually infected.
Precision = 8/10 == 80%

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

PERFORMANCE EVALUATION METRICS - PRECISION & RECALL

Let'sdescribethesetwoconceptsthroughwords&intuition

Recall of Class A:

Fromallthedatapointsthattrulywere"A",howmanydidwepredict?

Example: From a universe of 1000 people, 100 are actually infected with coronavirus(wedonotknowthis,ofcourse)wecorrectlypredictedthat50 peoplehadthevirus
Recall=50/100==50%

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

PERFORMANCE EVALUATION METRICS - PRECISION & RECALL

Which one should we prioritize?

DEPENDS ON THE PROBLEM

However, we have to learn how to decide which one we want to prioritize: Predict if a patient has cancer

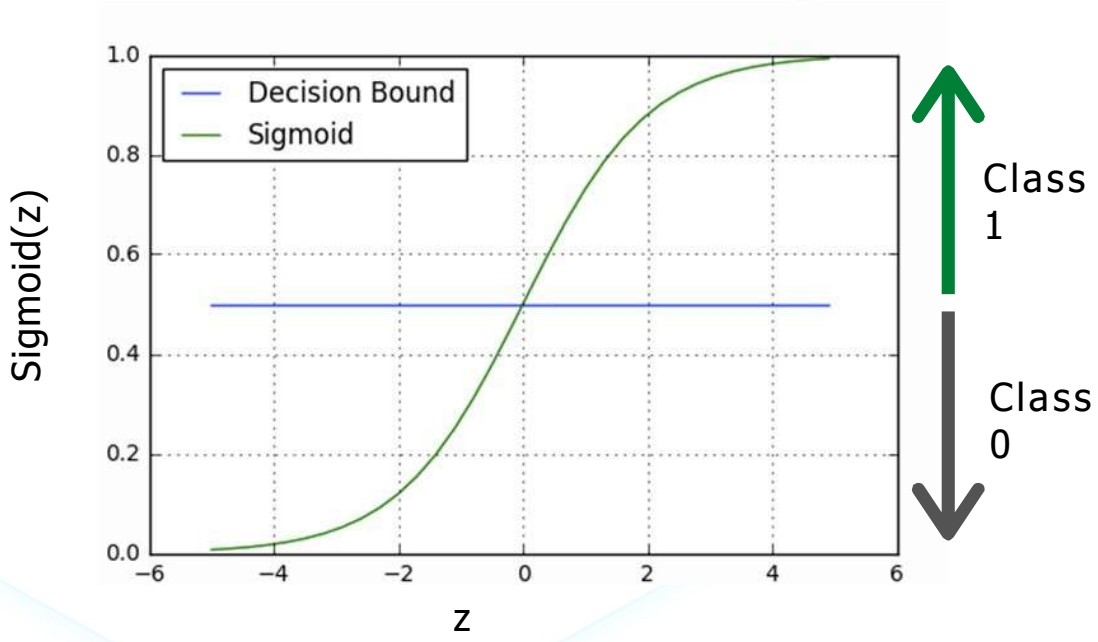
Brute Force Precision: only claim the patient has cancer if every single test comes up true multiple times in a row

Brute Force Recall: assume that a patient has cancer "to play it safe" before more tests come in

PERFORMANCE EVALUATION METRICS - PRECISION & RECALL

Which one should we prioritize?

PERFORMANCE EVALUATION METRICS - LOGISTIC REGRESSION



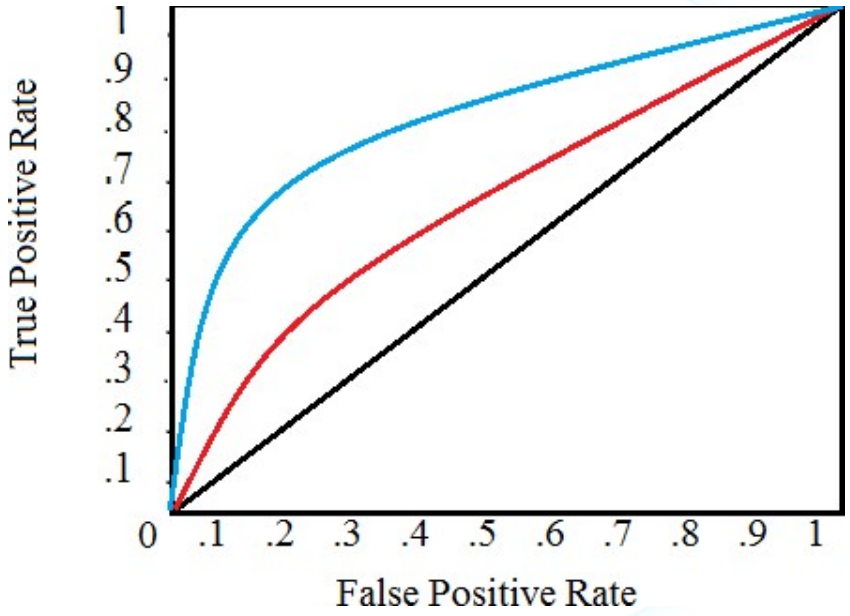
How do you relate the decision boundary to precision/recall in the logistic regression algorithm?

PERFORMANCE EVALUATION METRICS - F1 SCORE

A Trade-off between both such that the algorithm doesn't improve too much one of them at the expense of the other

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

PERFORMANCE EVALUATION METRICS - ROC CURVE



It is a plot of the false positive rate (x-axis) versus the true positive rate (y-axis) for a number of different candidate threshold values between 0.0 and 1.0. Put another way, it plots the false alarm rate versus the hit rate.

ANY QUESTIONS ?