Logistic Regression

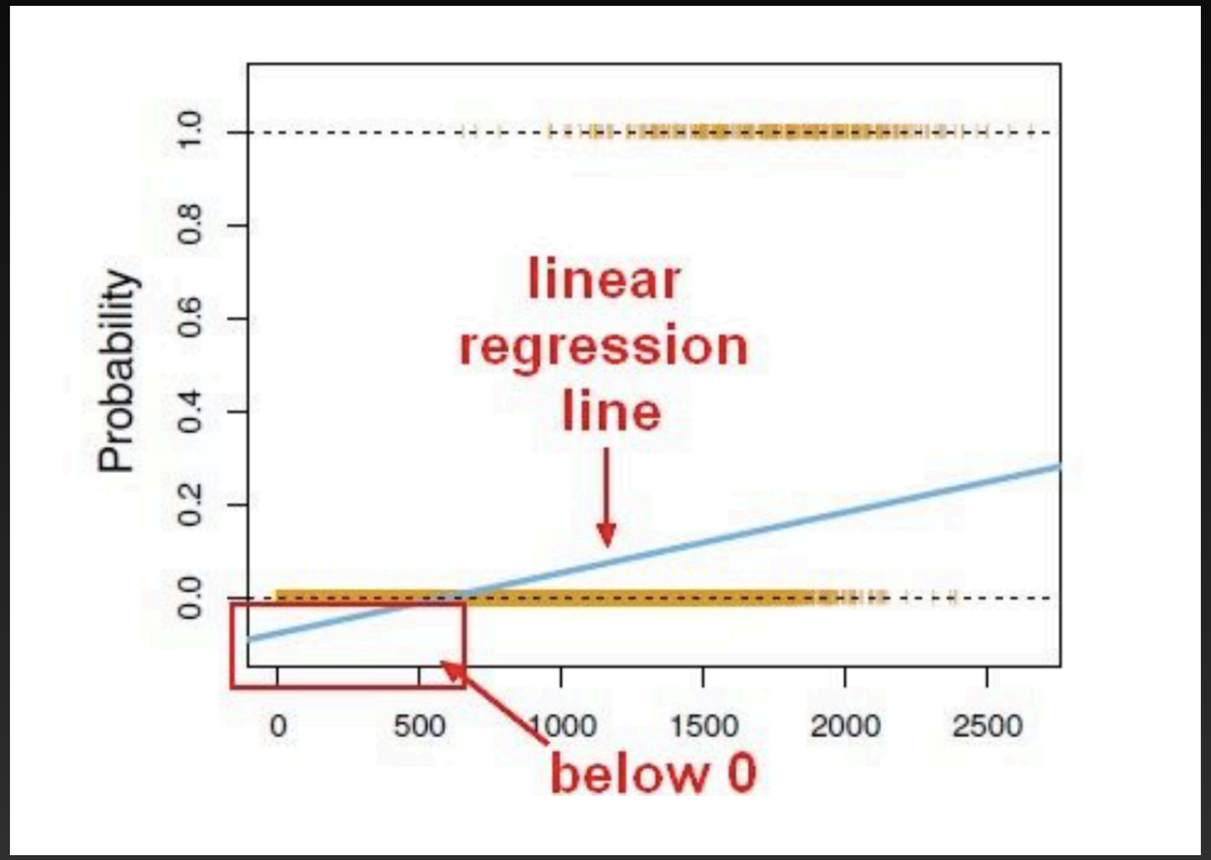
Logistic Regression for Classification

- Spam vs Ham Email
- Loan Default (Yes/No)
- Disease Diagnosis

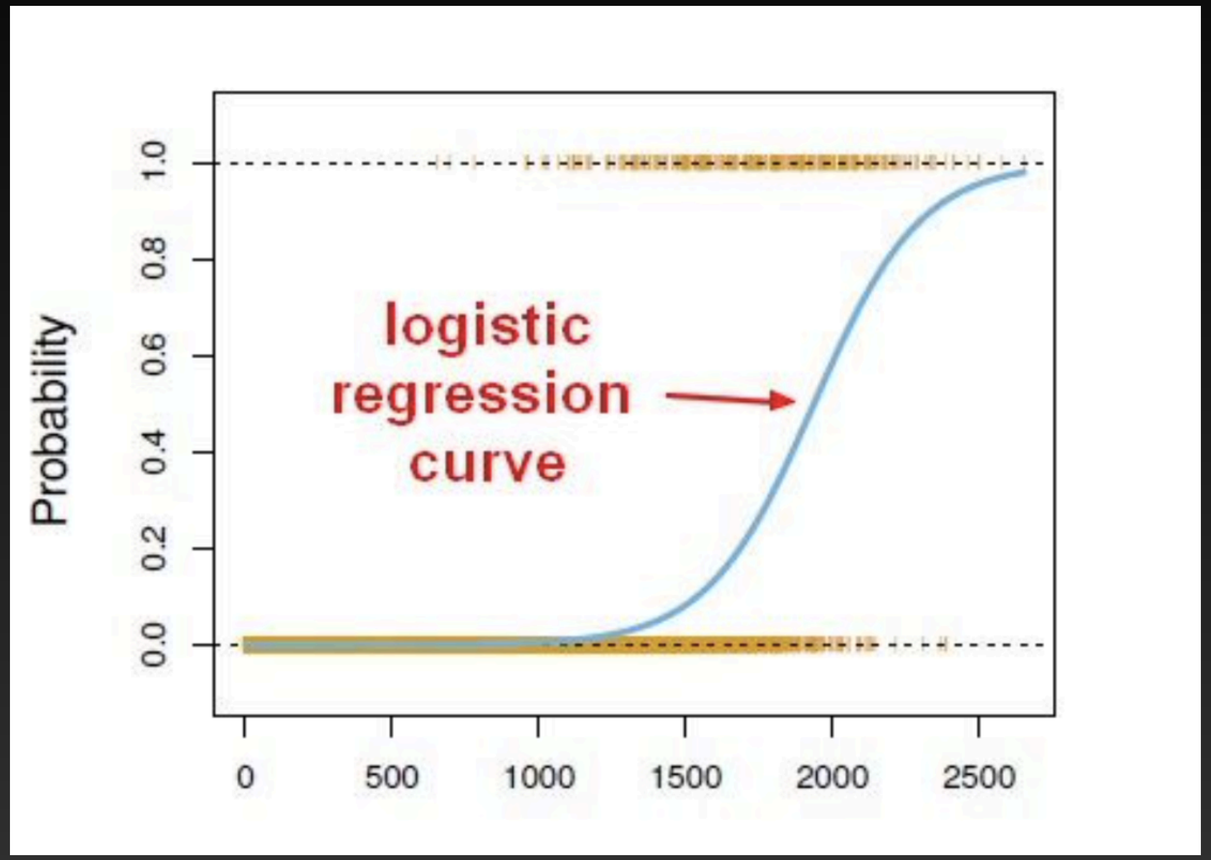
These were all examples of binary classification

The convention for binary classification is to have two classes: 0 & 1

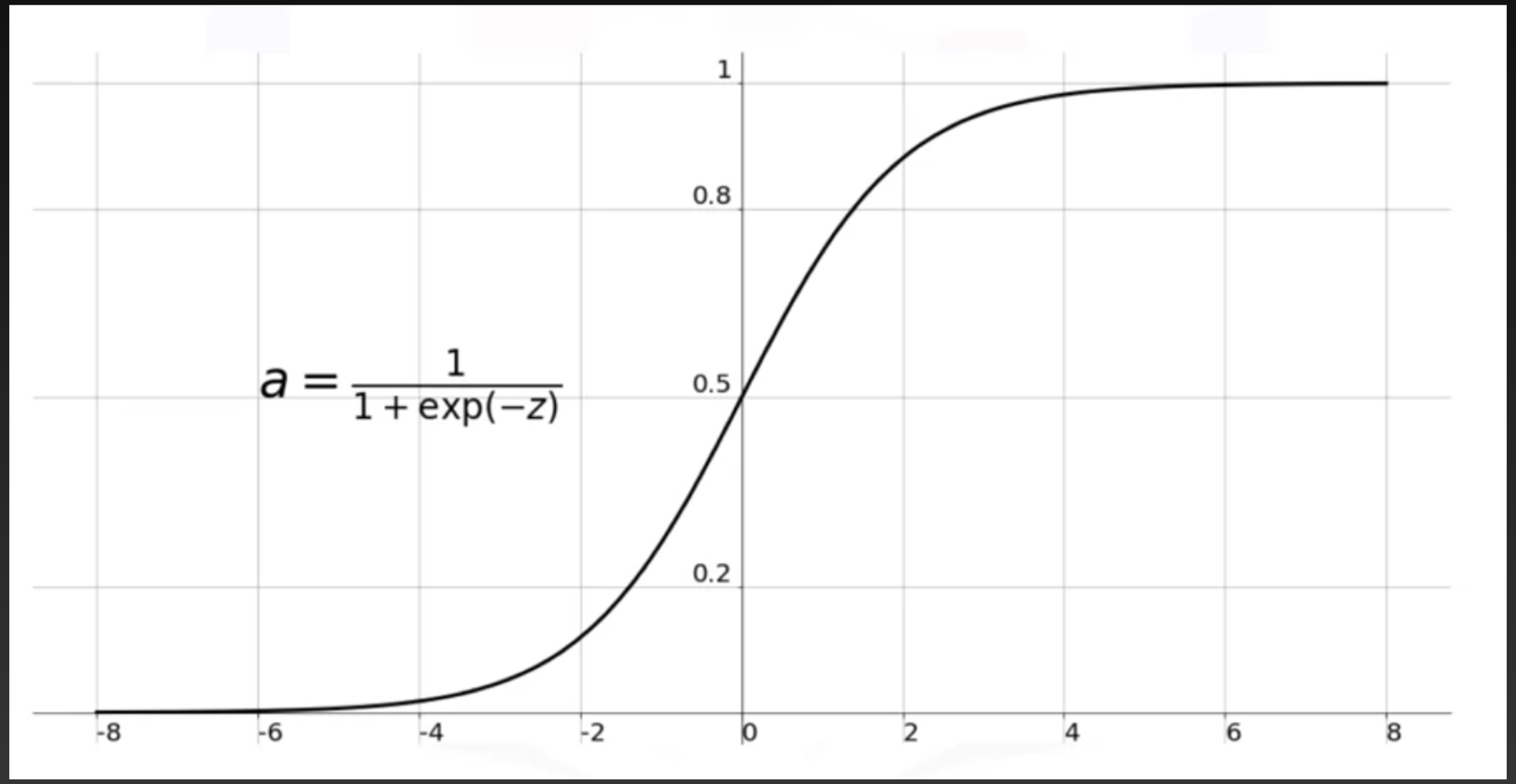
We can’t use a linear regression model for binary groups as it won’t be a good fit

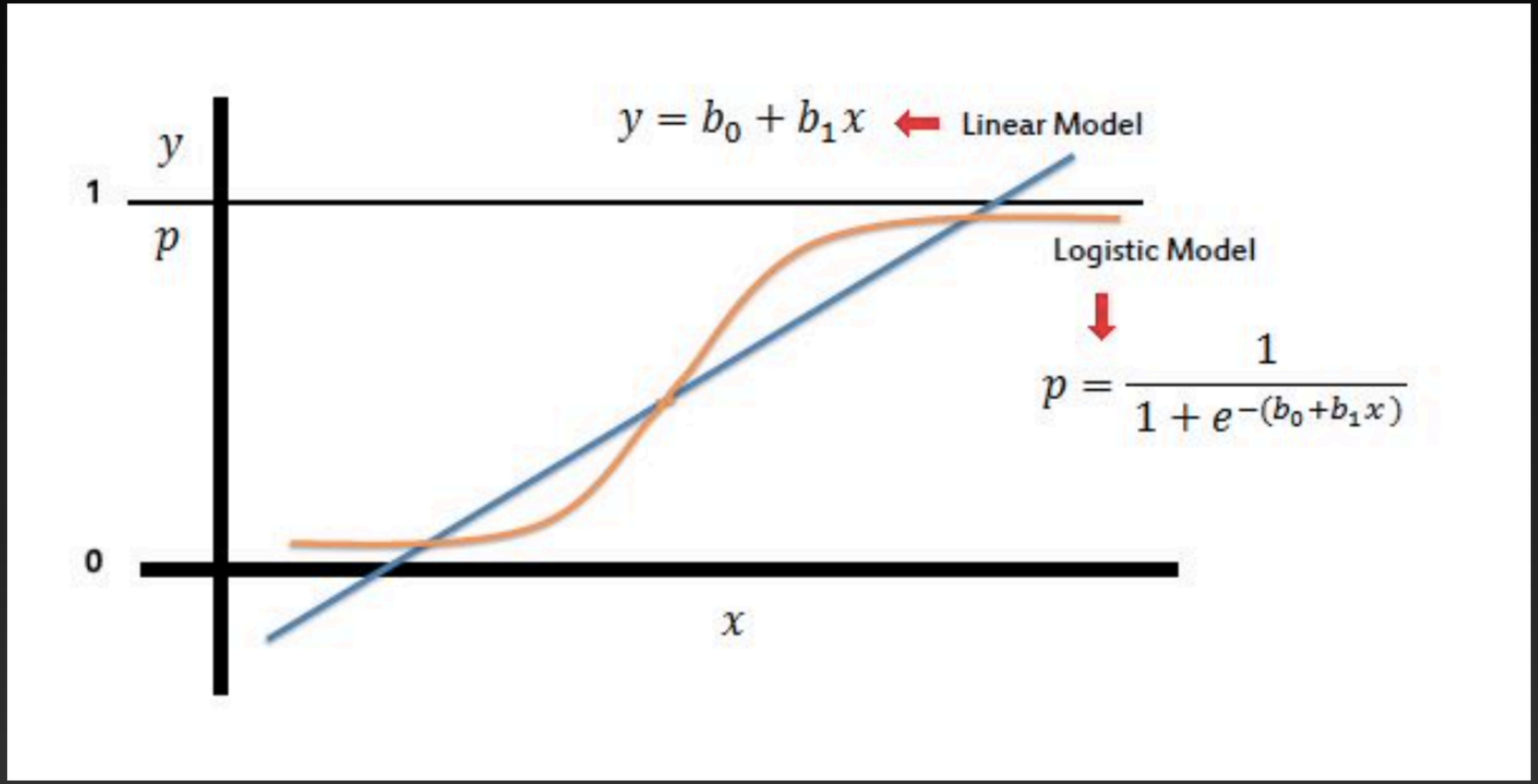


Instead, we can transform our linear regression line into a logistic regression curve



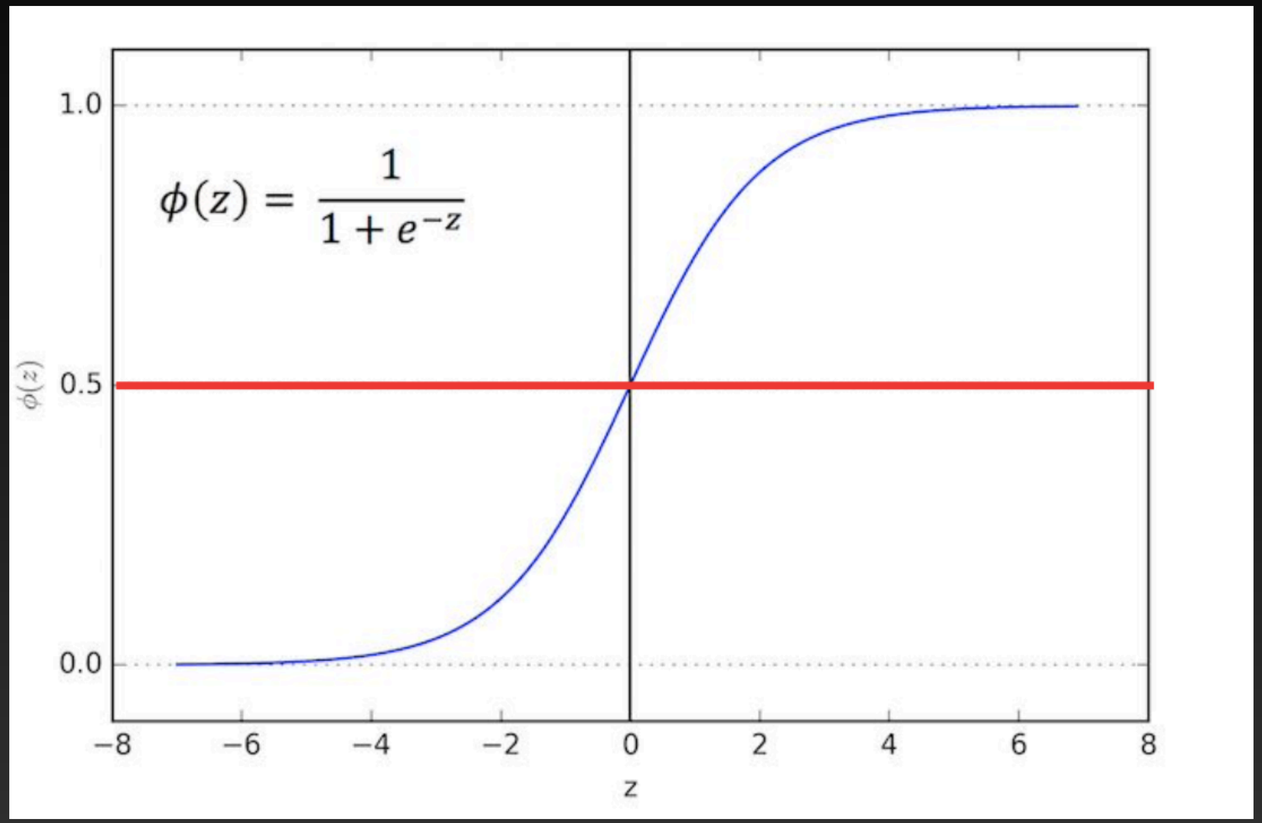
The Sigmoid Function takes in any input and outputs a value between 0 and 1




$y = b_0 + b_1 x$
Linear Model
y
1
p
Logistic Model
$p = \frac{1}{1 + e^{-(b_0 + b_1 x)}}$
0
x

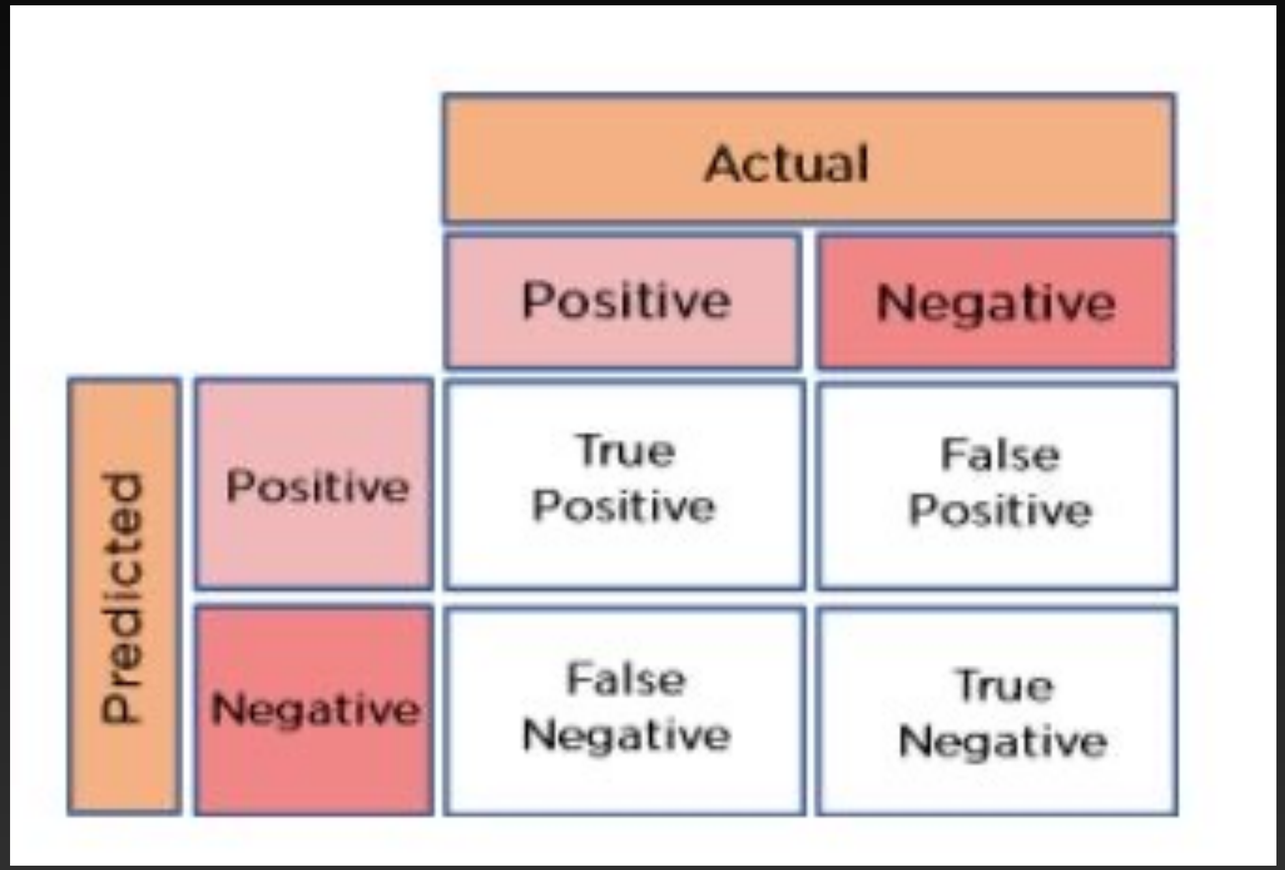
We can set a cutoff at 0.5, anything below it belongs in class 0 and above in class 1

Using the sigmoid function, we output a value between 0 and 1 and classify our data



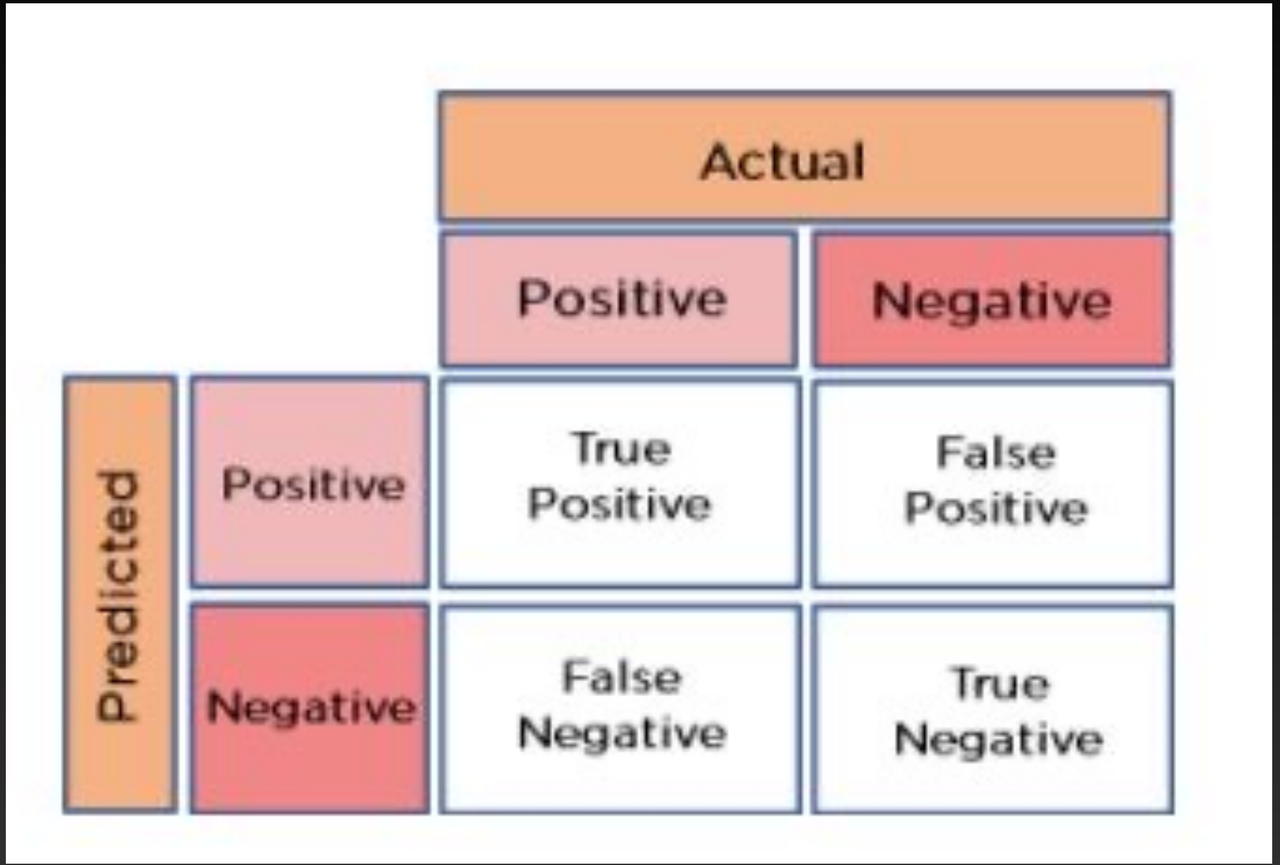
Error Metrics

Confusion Matrix

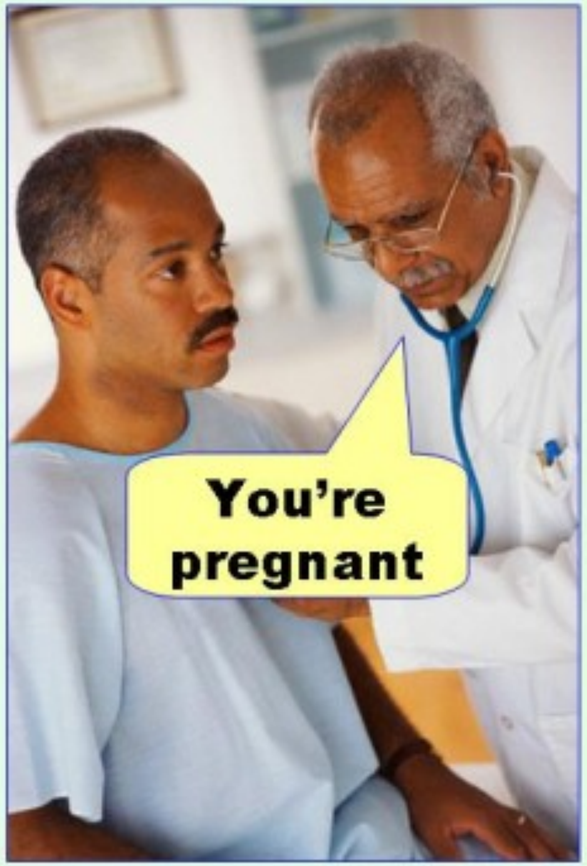


Basic Terminology:

- True Positive (TP)
- True Negative (TN)
- False Positive (FP)
- False Negative (FN)



Type I error
(false positive)



Type II error
(false negative)