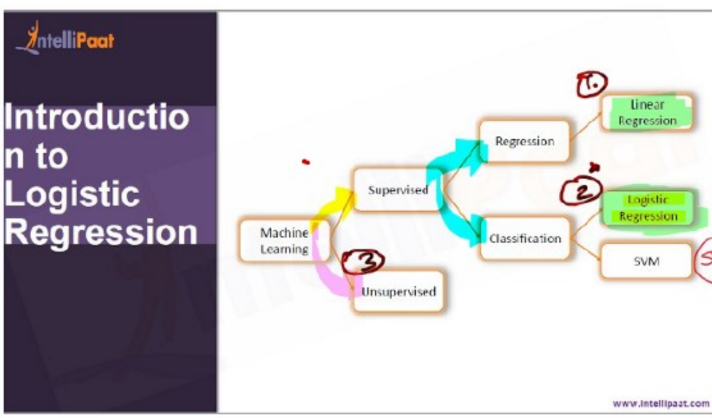
Logistic Regression

02 March 2024 20:04

Logistic vs Logistics

Logistic → ML algorithm
- supervised learning (classification)

Logistics → Key component of supply chain Management.
- it means transportation which can be done in different modes
  - Rail – RMS
  - Road – VRL
  - Air – DHL FedEx
  - Ocean – Suez Canal.



(Support Vector Machines)

Whether it will rain Tomorrow or not?

Patient tumour is → Benign
→ Malign

* [Econometrics]

Logistic regression is a statistical model / machine learning algorithm that uses a logistic function (logit) to model a binary (categorical) dependent variable.

2 classes – Yes
– No

ON 0 Spam

OFF 1 Ham (Non-Spam)

$$Y = f(x_1, x_2, x_3, \dots x_n)$$

(Target variable

input | predictor | independent | features

Categorical { Target variable / Response / Dependent

input | predictor | independent | features → variables

either continuous or categorical

Logistic Regression

$\rho / r \rightarrow$

$0.7 \le r/\rho \le 1 \rightarrow$ $\boxed{\text{strong correlation}}$

$0.3 \le r/\rho < 0.7 \rightarrow$ Moderate "

→ weak correlation

$p \rightarrow \ge 0.5 \rightarrow$ TRUE

$p \rightarrow 0 < p < 0.5 \rightarrow$ False.

$\boxed{0 \le P \le 1}$

y →

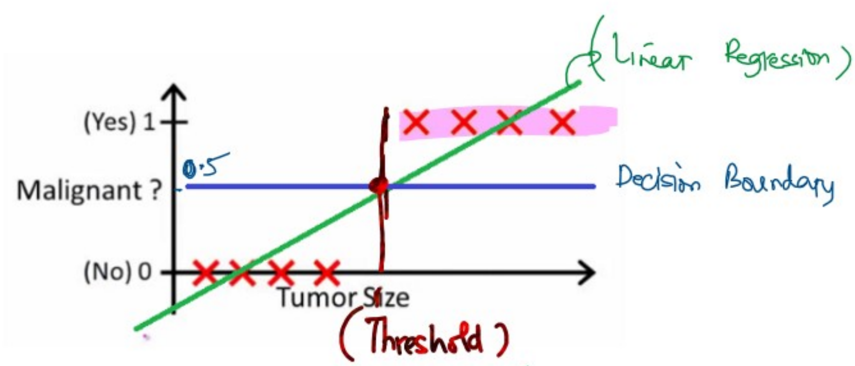
0.315
0.157
0.715
0.658
0.90

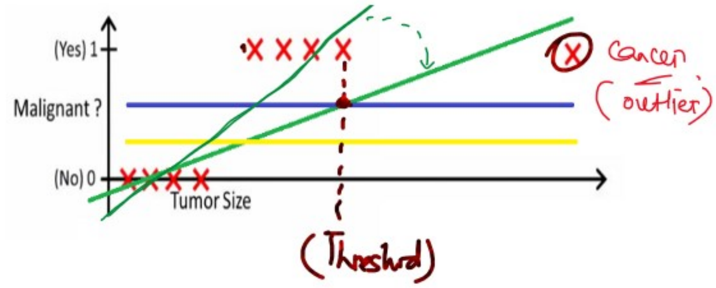
→ < 0.5 → False ; ≥ 0.5 → True

(0 to 1)

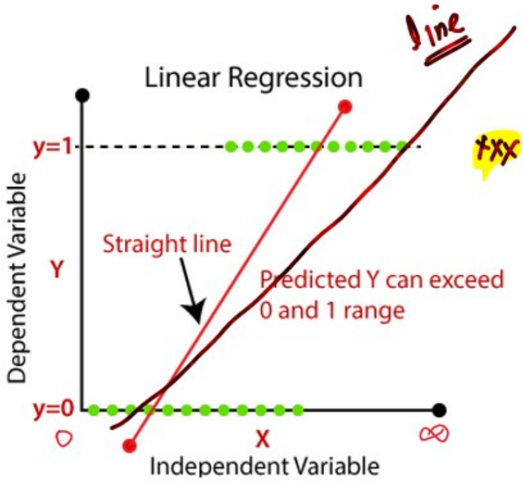
y	Y
0.315	F
0.157	F
0.715	T
0.658	T
0.90	T

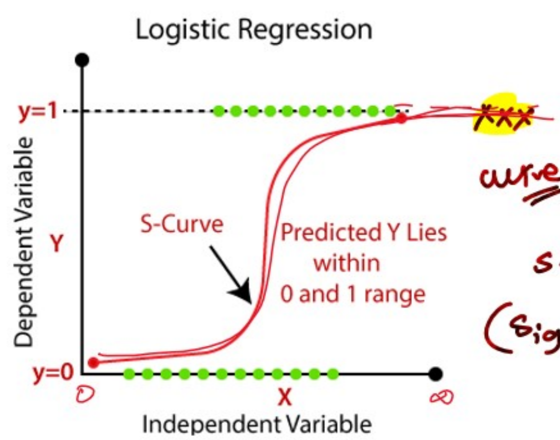
Why do we need a classifier?





xxx → more data points
likely outliers





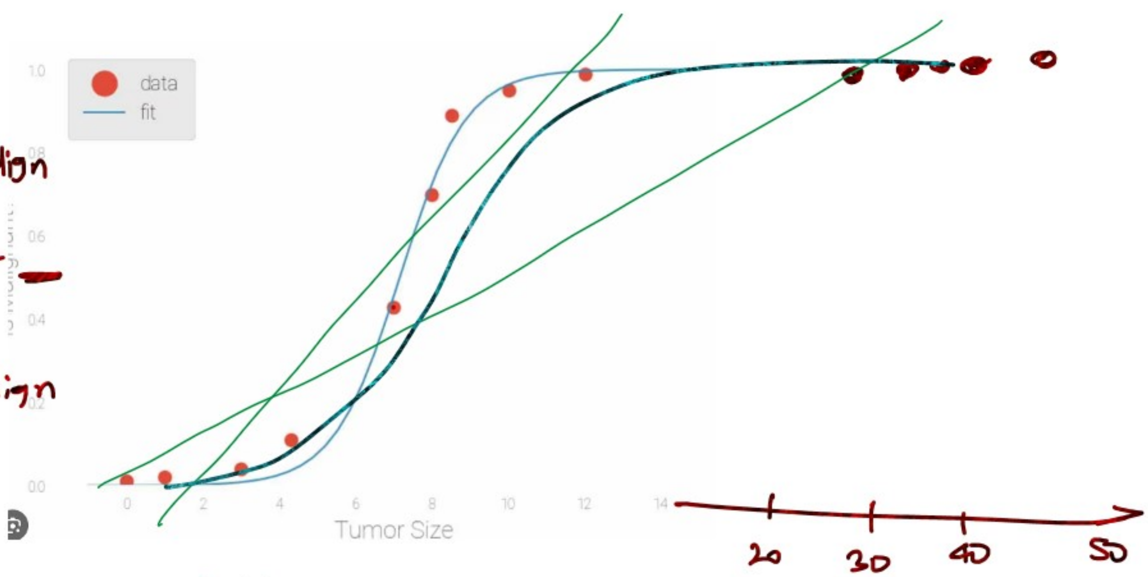
curve.

s-curve
(Sigmoid)

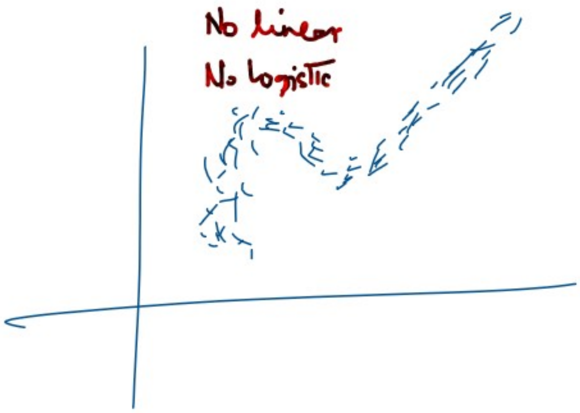
Malign

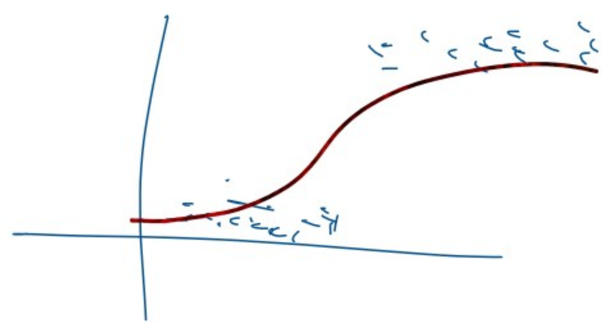
0.5

Benign



No linear
No logistic





Mathematics for logistic regression

Sigmoid Function

$$f(x) = \boxed{\left(\frac{1}{1+e^{-x}}\right)}$$

$g(x) = e^{-x}$ e: Napier's constant

$\downarrow$ $e \simeq 2.71$

exponential decaying

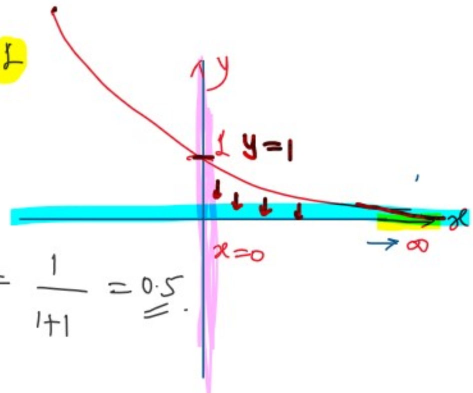
$e \simeq 2.71$

$e^{-\infty} = \frac{1}{e^{\infty}} = \frac{1}{\infty} \to 0.$

$x \to \infty$ $f(x) = \lim_{x\to\infty} \left(\frac{1}{1+e^{-x}}\right)$

$$= \frac{1}{(1+e^{-\infty})} = \frac{1}{1+0} = 1$$

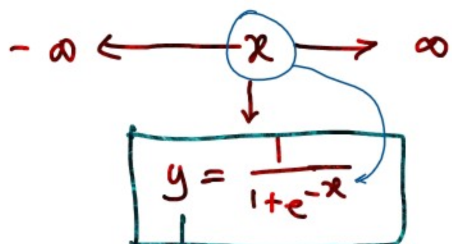
$e^{-\infty} \to 0.$



$e^{-x} \to x \to \infty \quad e^{-x} \to 0$

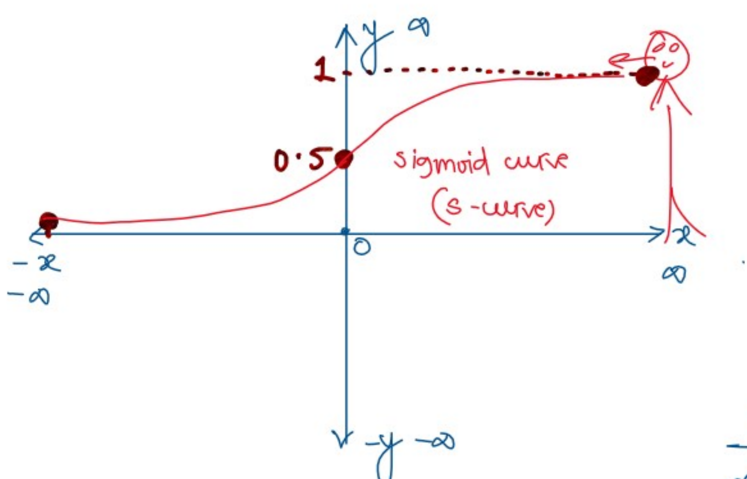
$x = 0$ $\lim_{x\to 0}\left(\frac{1}{1+e^{-x}}\right) = \left(\frac{1}{1+e^{-0}}\right) = \frac{1}{1+1} = 0.5$

$x \to -\infty$ $\lim_{x\to\infty}\left(\frac{1}{1+e^{-x}}\right) = \frac{1}{\left(1+e^{-(-\infty)}\right)} = \left(\frac{1}{1+e^{\infty}}\right) \to \infty. = \frac{1}{1+\infty} = \frac{1}{\infty} = 0.$



$$y = \frac{1}{1+e^{-x}}$$

0 to 1



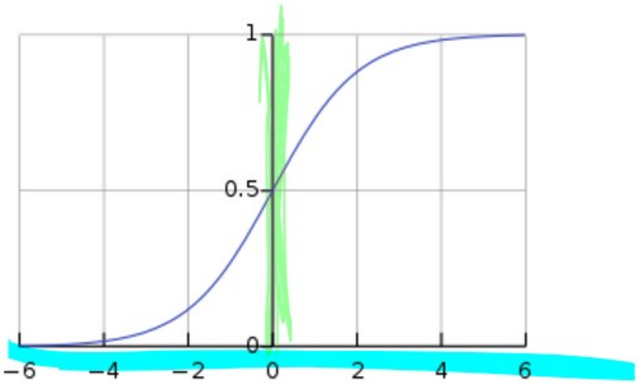
Input variable

$x_1 \mid x_2$

social Media

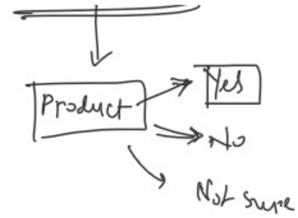
Yes

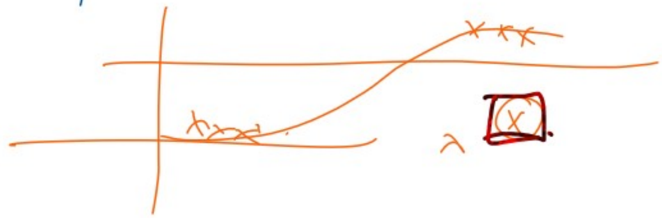
$x - y - \infty$



Input variable

x_1	x_2
0 to ∞	$-\infty$ to ∞.





Sigmoid function

30 languages

Article Talk Read Edit View history Tools

From Wikipedia, the free encyclopedia

A **sigmoid function** is any mathematical function whose graph has a characteristic S-shaped curve or **sigmoid curve**.

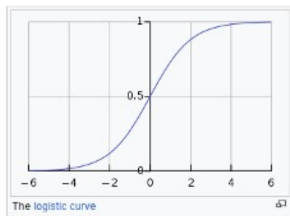
A common example of a sigmoid function is the logistic function shown in the first figure and defined by the formula:[1]

$$\sigma(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{1+e^x} = 1 - \sigma(-x).$$

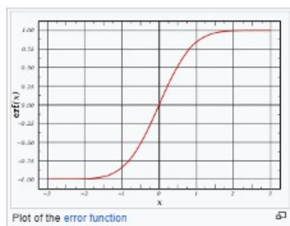
Other standard sigmoid functions are given in the Examples section. In some fields, most notably in the context of artificial neural networks, the term "sigmoid function" is used as an alias for the logistic function.

Special cases of the sigmoid function include the Gompertz curve (used in modeling systems that saturate at large values of x) and the ogee curve (used in the spillway of some dams). Sigmoid functions have domain of all real numbers, with return (response) value commonly monotonically increasing but could be decreasing. Sigmoid functions most often show a return value (y axis) in the range 0 to 1. Another commonly used range is from −1 to 1.

A wide variety of sigmoid functions including the logistic and hyperbolic tangent functions have been used as the activation function of artificial neurons. Sigmoid curves are also common in statistics as cumulative distribution functions (which go from 0 to 1), such as the integrals of the logistic density, the normal density, and Student's *t* probability density functions. The logistic sigmoid function is invertible, and its inverse is the logit function.



The logistic curve



Plot of the error function

Definition [edit]

A sigmoid function is a bounded, differentiable, real function that is defined for all real input values and has a non-negative derivative at each point[1][2] and exactly one inflection point.

https://en.wikipedia.org/wiki/Sigmoid_function

$$y = f(x) = \frac{1}{1+e^{-x}}$$

$(\beta_0 + \beta_1 x)$

linear regression.

$-\infty < x < \infty$

$$\boxed{p = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}}}$$

Logit or logistic function

Log (odds)

What do you mean by odds?

$$\frac{p}{(1-p)} = \text{Odds}$$

odds in favour of an event:

$$\left(\frac{\text{probability of the event}}{\text{probability of non-event}}\right)$$

let us say probability of winning a game is 0.6.

odds in favour of winning $\Rightarrow \left(\frac{p}{1-p}\right) = \left(\frac{0.6}{1-0.6}\right) = \frac{0.6}{0.4} = \frac{6}{4} = 1.5$

For every 1.5 successes, there is 1 failure.

For every 3 successes, " 2 failures.

$$p = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}} \quad (\beta_0+\beta_1 x) \rightarrow y.$$

$$p = \frac{1}{1+e^{-y}} \text{ ——— (1)}$$

$$(1-p) = \left(1 - \frac{1}{1+e^{-y}}\right)$$

$$(1-p) = \left(\frac{\cancel{1}+e^{-y}-\cancel{1}}{1+e^{-y}}\right) = \left(\frac{e^{-y}}{1+e^{-y}}\right) \text{ ——— (2)}$$

$$\left(\frac{p}{1-p}\right) = \frac{\left(\frac{1}{\cancel{1+e^{-y}}}\right)}{\left(\frac{e^{-y}}{\cancel{1+e^{-y}}}\right)} = \frac{1}{e^{-y}}$$

$$\left(\frac{p}{1-p}\right) = \frac{\cancel{\left(1+e^{-y}\right)}}{\left(\frac{e^{-y}}{\cancel{1+e^{-y}}}\right)} = \frac{1}{e^{-y}}$$

$$\left(\frac{p}{1-p}\right) = \frac{1}{e^{-(\beta_0+\beta_1 x)}}$$

$$\left(\frac{p}{1-p}\right) = e^{(\beta_0+\beta_1 x)}$$

$$\log_e\left(\frac{p}{1-p}\right) = \log_e e^{(\beta_0+\beta_1 x)} = (\beta_0+\beta_1 x)$$

$$\boxed{\underbrace{\log_e\left(\frac{p}{1-p}\right)}_{\log(\text{odds})} = (\beta_0+\beta_1 x)} \quad *$$

logit function

↓

odds of 'p'

↓

log(odds)

In general,

$$\underbrace{\underbrace{\log_e(\text{odds})}_{\log_e\left(\frac{p}{1-p}\right)}}_{\downarrow\ o/p} = \underbrace{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_n x_n}_{\text{input}}.$$

\# log(odds) represents the natural log of the odds of the event happening.

\# $\beta_0, \beta_1, \beta_2 \ldots \beta_n$ are the coefficients associated with each predictor $X_1, X_2, \ldots X_n$ respectively

intercept (β_0)

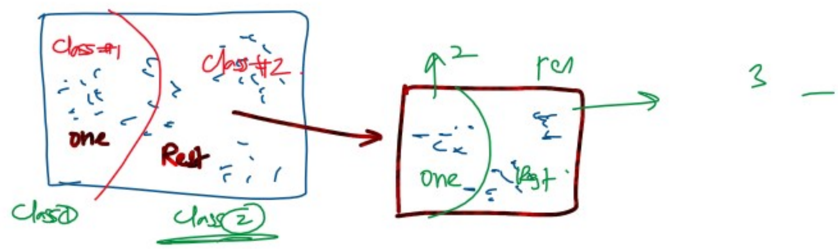
$$\beta_1 x_1 \rightarrow +\beta_1 x_1 \Rightarrow \log(\text{odds}) \uparrow\uparrow\uparrow$$

or

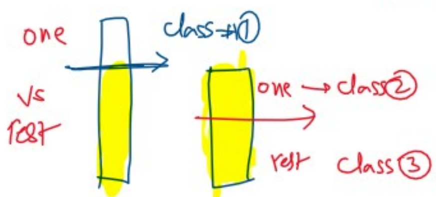
$$\beta_1 x_1 \rightarrow -\beta_1 x_1 \Rightarrow \log(\text{odds}) \downarrow\downarrow\downarrow$$

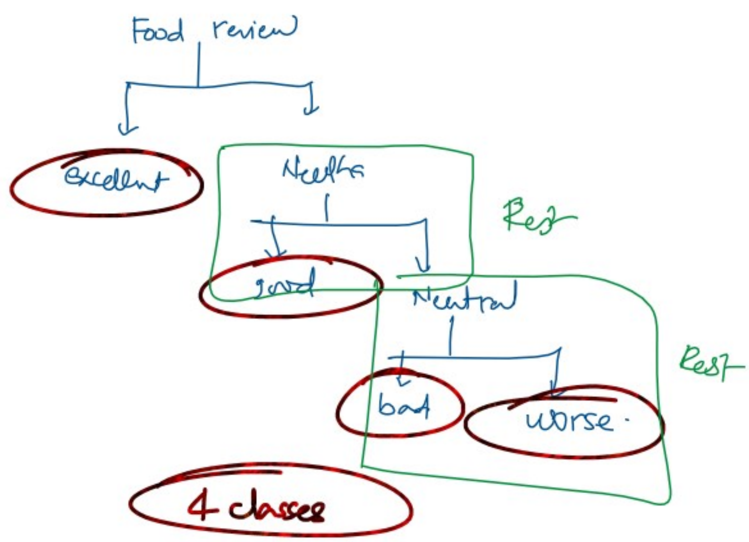
Logistic Regression for multiclass problems

* One vs Rest (OVR)



4 class.

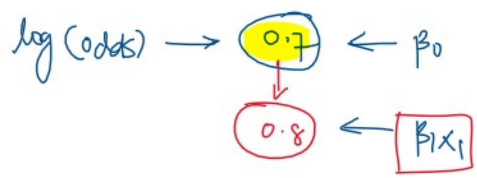




What is the purpose of intercept in logistic?

- all the predictor variables ($x_1, x_2 \dots x_n$) are set to zero, intercept (β_0) provides a baseline level of log of odds.

- to get baseline log of odds



Assumptions of Logistic Regression

① Response | Target variable is categorical
(binary)
or
multiclass

② Predictor (input) variables are independent.

(2) Predictor (input) variables are independent.

y	x_1	x_2	x_3

$x_2 = x_1 2$

multicollinearity

- Observations should not come from repeated measurements of the same variable

(3) Sample size is sufficiently large.

(4) No extreme outliers.

(5) Linear relationship between input variables and logit of the response variables.

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

Evaluating the logistic regression model

Confusion Matrix

It is a performance measurement for machine learning classification problem where output can be two or more classes.

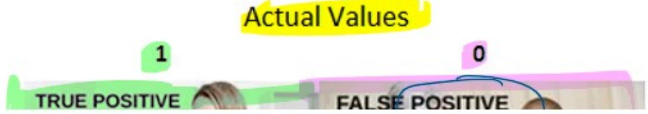
Binary class

Actual Values

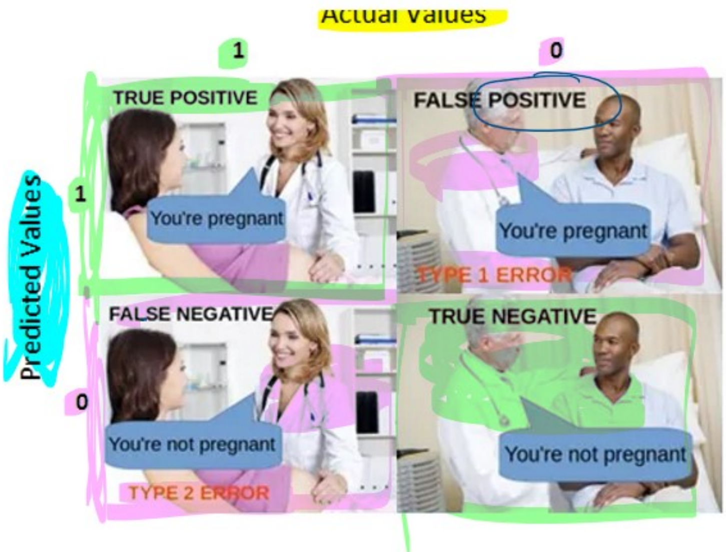
Predicted Values	Positive Class (1)	Negative (0) Class
Positive Class (1)	TP: True (correct) Positive (class)	FP: False Positive (Type 1 error)
Negative Class (0)	FN: False Negative (Type 2 error)	TN: True Negative

2x2 Matrix

Actual Values



Positive → Pregnant



Positive → Pregnant
Negative → Not Pregnant

Agenda:

1. Multicollinearity & VIF
2. Performance Metrics - Accuracy, Precision ..blah -- Imbalanced Data
3. Interpreting Linear and Logistic coeff - Statmodels
4. Pros/Cons for Linear and Log
5. Decision Trees - Gentle introduction

Accuracy

			Actual (True) Values	
			Cancer	No Cancer
Positive	Predicted Values	Cancer	45	18
Negative	Predicted Values	No Cancer	12	25

Confusion matrix for the cancer example. Image by Author.

$$\text{Accuracy} = ?\ \frac{45+25}{45+18+12+25}$$

$$= \frac{70}{\cancel{100}} \times \cancel{100} = 70\%$$

True Positive : 45

True Negative 25

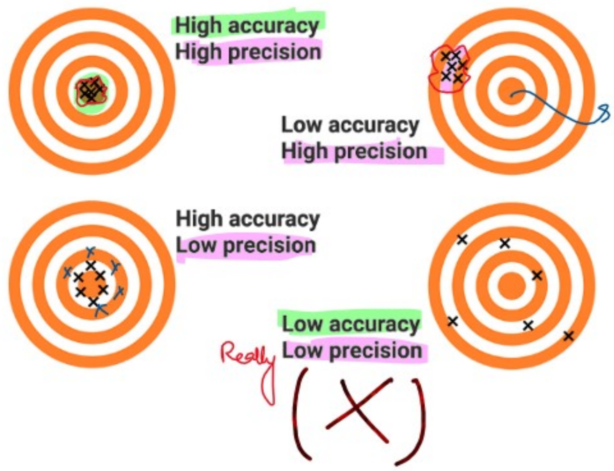
False Positive : 18

False Negative : 12

What's the difference between Accuracy and Precision ?

Accuracy vs Precision.

Accuracy vs Precision.



Accuracy (close to the target)

Accurately hitting the target (bull's eye) implies you are close to the center of the target.

(Bull's eye)

Precision (consistency)

Precisely hitting a target where all the hits are closely spaced even if they are really far from the center of the target.

Precision is a measure of how many of the positive predictions made are correct.

$$\text{Precision} = \left(\frac{TP}{TP + FP}\right) = \left(\frac{\text{No. of correctly predicted positive classes}}{\text{No. of total positive classes}}\right)$$

$$= \frac{45}{45+18} = \frac{45}{63} \times 100 = 71.4\%$$

True Positive : 45

True Negative 25

False Positive : 18

False Negative : 12

Recall | Sensitivity

Recall | sensitivity is a measure of how many of the positive cases, the classifier correctly predicted, over all the positive cases (actual) in the data.

$$\text{Recall/Sensitivity} = \left(\frac{Tp}{}\right) = \frac{45}{57} \times 100 = 78.9\%$$

(actual)

$$\text{Recall/Sensitivity} = \left(\frac{TP}{TP+FN}\right) = \frac{45}{57} \times 100 = 78.9\%$$

Specificity

It is a measure of how many negative predictions made are correct (True Negatives)

$$\text{specificity} = \left(\frac{TN}{TN+FP}\right) = \frac{25}{25+18} \times 100 = 58.1\%$$

Model outcomes

1. Accurate ≙ 95% Precision ≙ 30%
2. Accurate ≙ 75% Precision ≙ 90%

Trainer #1 every topic → 90% P 40%

#2 → 70-75% P 90%

Recall | Sensitivity

[Medical diagnosis / Fraud detection tasks]

		Actual (True) Values	
		Cancer	No Cancer
Predicted Values	Cancer	45	18
	No Cancer	12	25

Confusion matrix for the cancer example. Image by Author.

Recall = 78.9% ↑↑ (99.9%)

$45+12 = 57$ had cancer

↓

out of 57, 12 guys got wrong diagnosis → we missed to detect cancer in the early stage.

also had cancer

↓

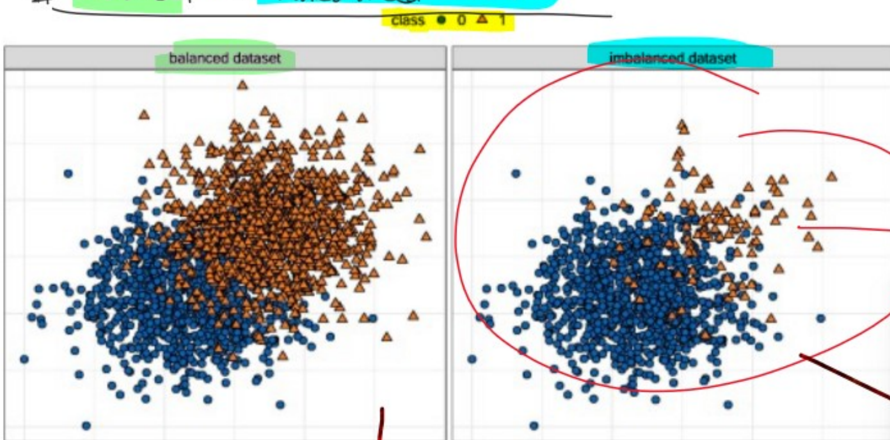
(but we missed)

F1-score:

It is a measure combining both precision and recall. It is expressed as the harmonic mean of precision and recall.

$$F1\ score = \frac{2 * (Precision \times Recall)}{[Precision + Recall]}$$

Balanced vs Imbalanced Dataset



everything is blue ●

$accuracy = \frac{90}{100} \times 100 = 90\%$

(F1-score) Recall.

almost equal proportion for both the classes. use metric

(Accuracy)

10K rows

Class 0 → 6000

Class 1 → 4000

Class 0 → 90% | 80

Class 1 → 10% | 20

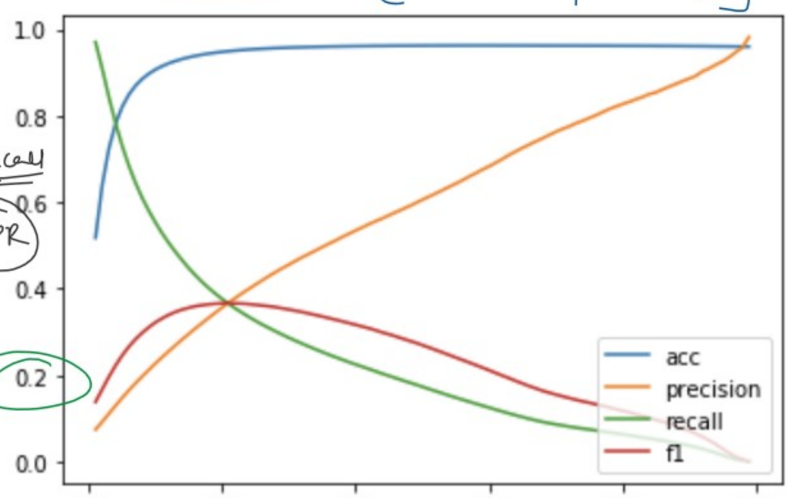
Note #. Accuracy is a good metric when the classes are balanced. However it may not be a suitable metric when there is a significant class imbalance because a model could achieve high accuracy by simply predicting the majority class.

Note: Precision, recall and F1 - score provide better Insights into the model's performance on positive instances which are often the minority class in the imbalanced datasets

Imbalanced dataset [4% class 1 | 96% class 0]

Recall

TPR



FPR → (1 - specificity) →

AUC

Positive class → 200

Negative class → $\boxed{1800}$

} 2K rows

1900 → Negative class (1800 N → N
100 P → N)

100 → Positive class

$$\text{Accuracy} = \frac{\cancel{1900}}{\cancel{2000}} \times 1\cancel{00} = 95\%$$

$$\text{Recall} = \frac{100}{200} \times 100 = 50\% \downarrow\downarrow$$

AUC - ROC curve:

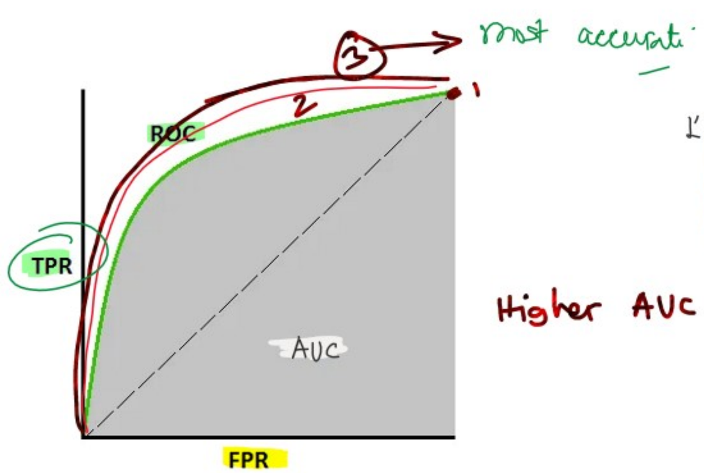
(Area under curve - Receiver Operating characteristics)

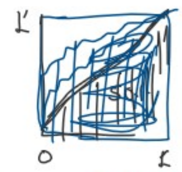
Y-axis: TPR : True Positive Rate (Recall)

X-axis: FPR : (1 - Specificity)

$$1 - \left(\frac{TN}{TN+FP}\right) = \left(\frac{\cancel{TN}+FP-\cancel{TN}}{TN+FP}\right) = \frac{FP}{TN+FP}$$

→ most accurate





Higher AUC → Higher Accuracy.