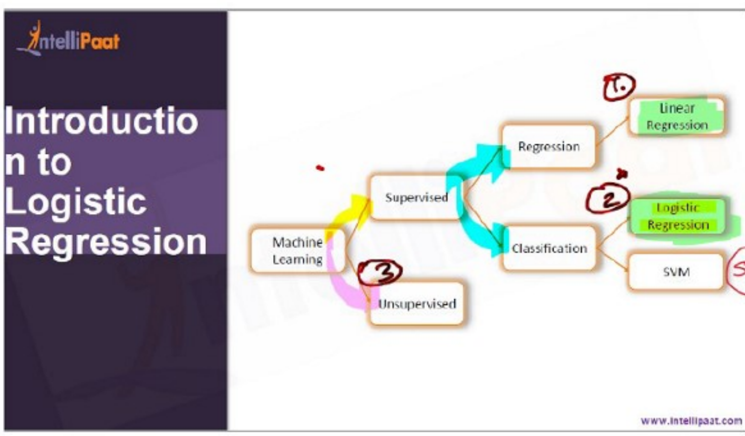
Logistic Regression

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)

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

input | predictor | independent | features
variables

Categorical
- Target variable
- Response
- Dependent

either continuous or categorical

Logistic Regression

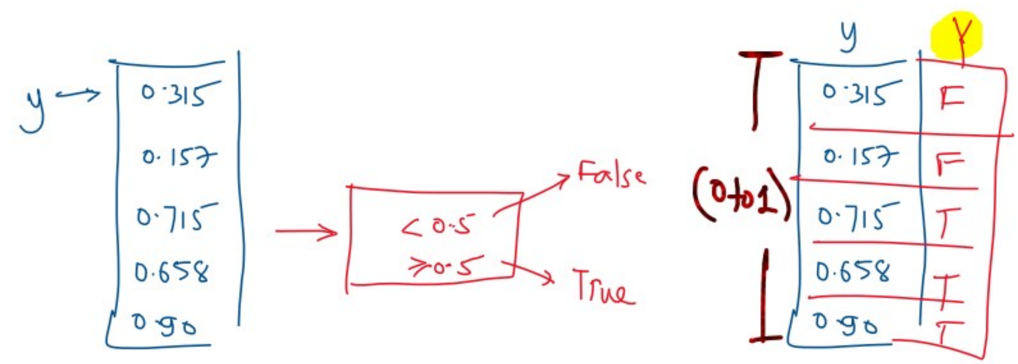
$\rho / r \rightarrow$

- $0.7 \le r|\rho \le 1 \rightarrow$ $\boxed{\text{strong correlation}}$
- $0.3 \le r|\rho < 0.7 \rightarrow$ Moderate "
- $\rightarrow$ weak correlation

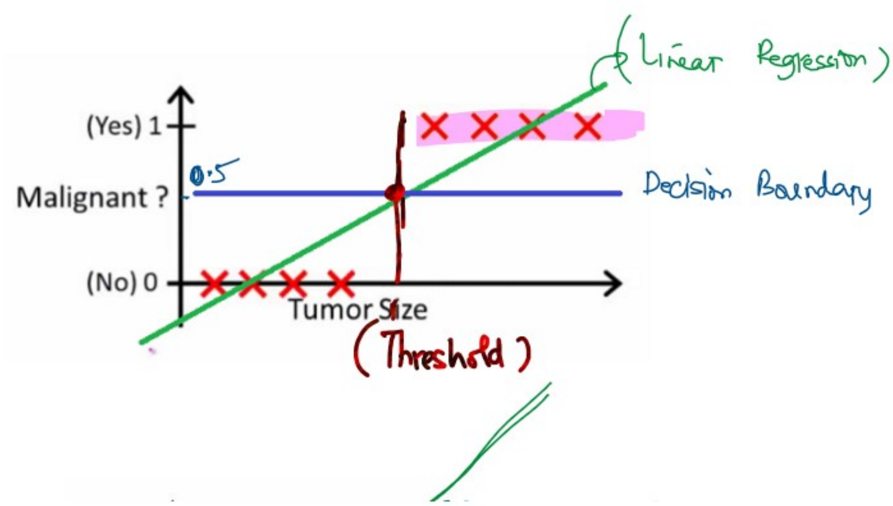
$p \rightarrow$

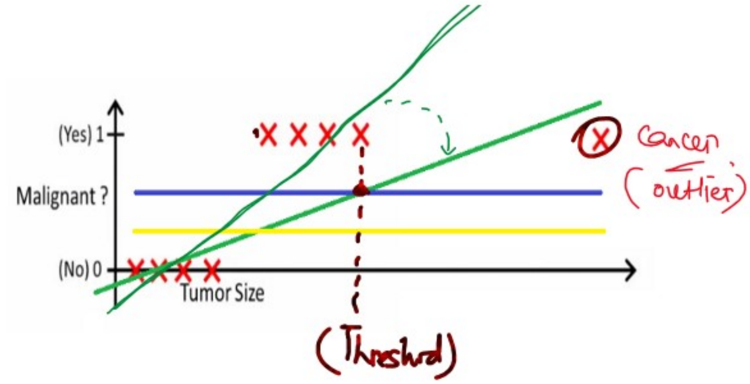
- $\ge 0.5 \rightarrow$ TRUE
- $0 < p < 0.5 \rightarrow$ False.

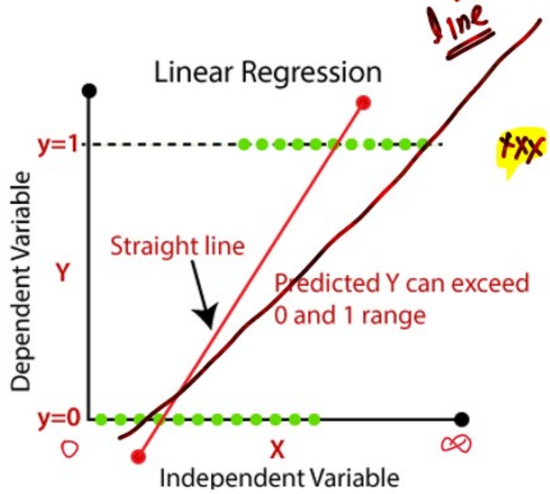
$\boxed{0 \le P \le 1}$

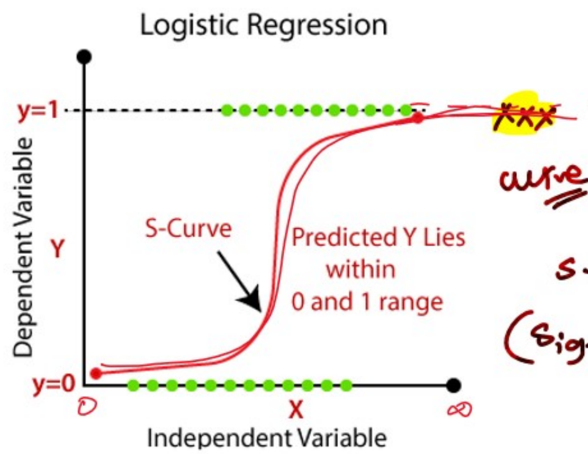


Why do we need a classifier?





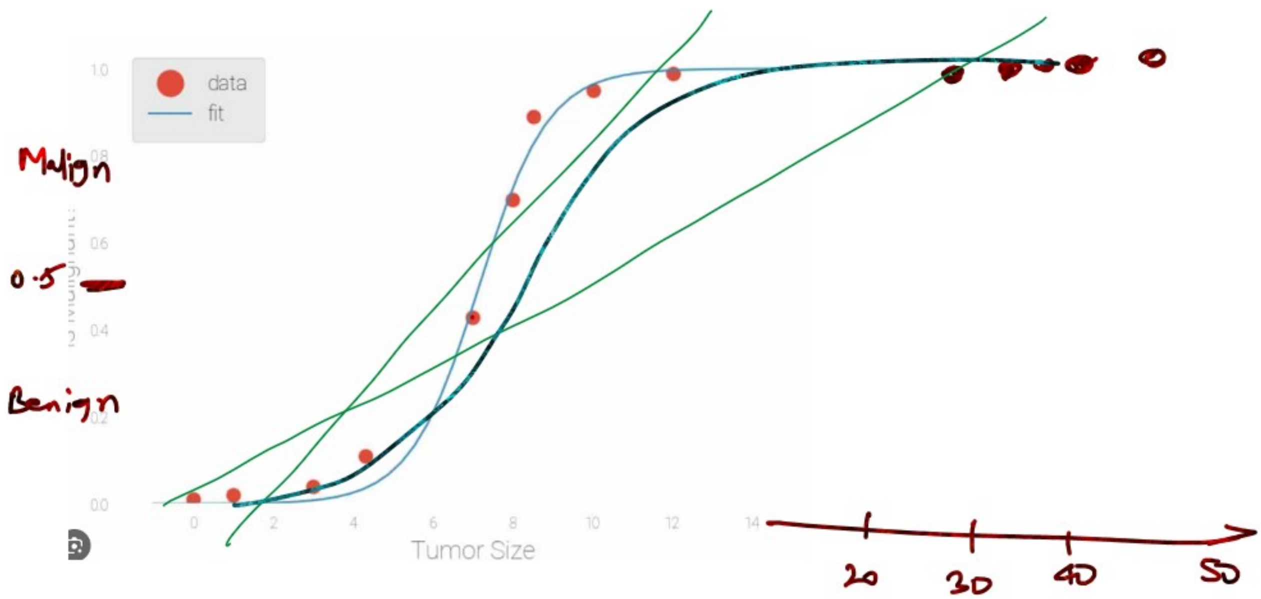


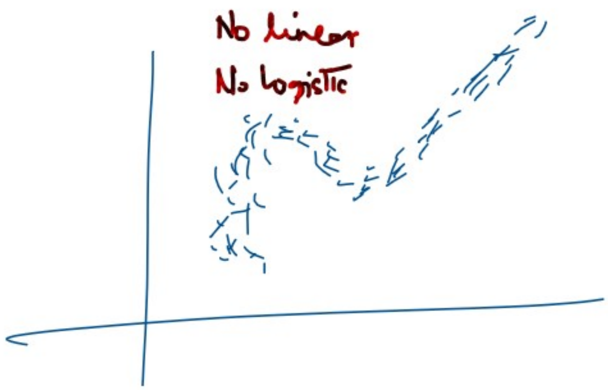


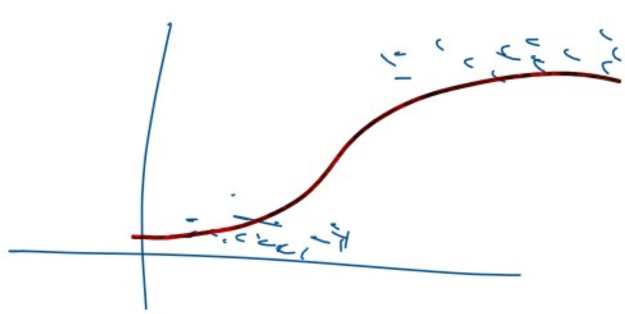
xxx → more data points
likely outliers

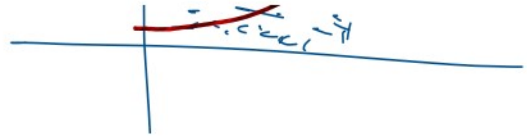
curve.

s-curve
(Sigmoid)









Mathematics for logistic regression

Sigmoid Function

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

$g(x) = e^{-x}$ $\downarrow$ exponential decaying

e : Napier's constant

$e \simeq 2.71$

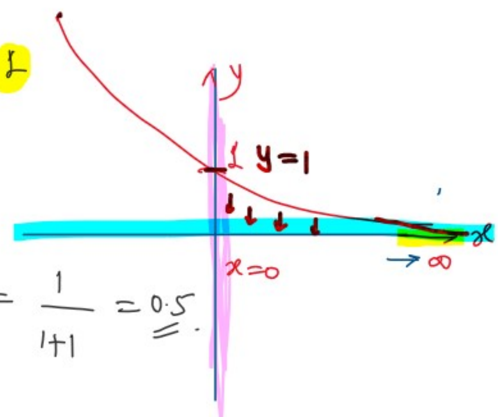
$e \simeq 2.71$

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

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

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

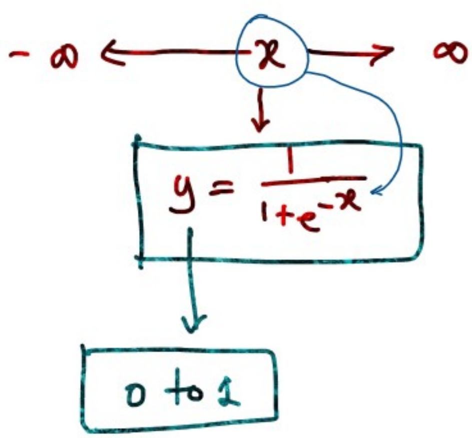
$e^{-\infty} \to 0.$

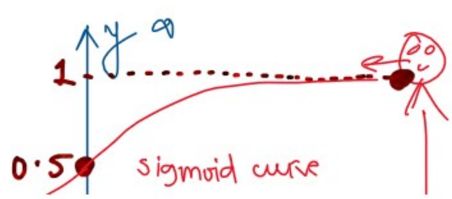


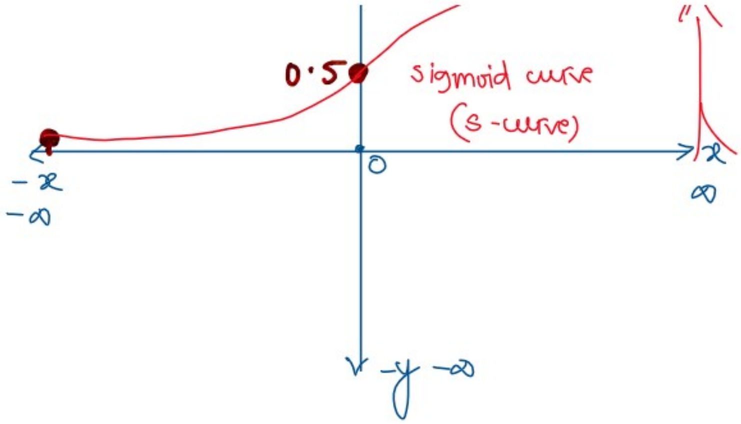
$e^{-x} \to x \to \infty \quad e^{-x} \to 0$

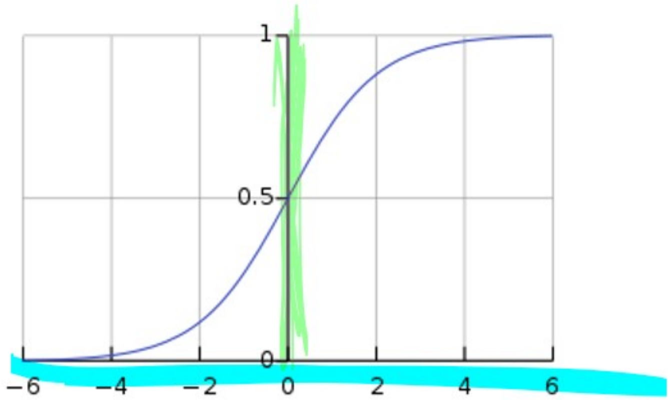
$x = 0$ $\quad \lim_{x\to 0}\left(\frac{1}{1+e^{-x}}\right) = \left(\frac{1}{1+e^{-0}}\right) = \frac{1}{1+1} = 0.5$

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



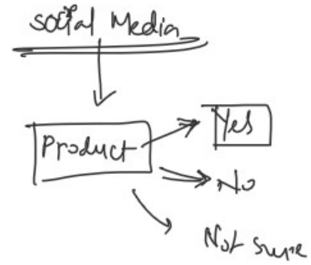


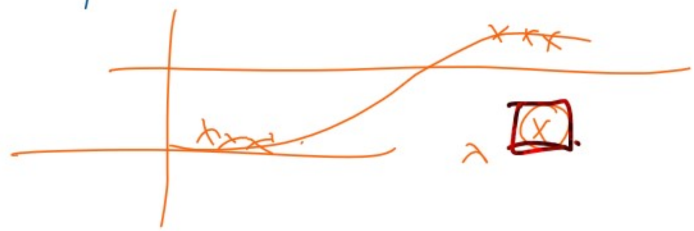




Input variable

x_1	x_2
0 to ∞	-∞ 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**.

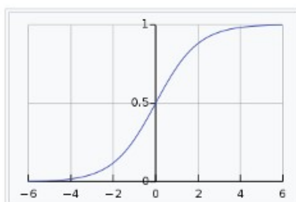
A common example of a sigmoid function is the logistic function shown in the first figure and defined by the formula:[1]

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

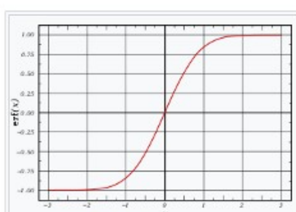
Other standard sigmoid functions are given in the Examples section. In some fields, most notably in the context of artificial neural networks, the term "sigmoid function" is used as an alias for the logistic function.

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.

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

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

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

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

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

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

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

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

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

logit function

↓

odds of 'p'

↓

log(odds)

In general:

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

$$\underbrace{\log_e\left(\frac{p}{1-p}\right)}_{o/p} \qquad \text{input}$$

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

$\beta_0, \beta_1, \beta_2 \cdots \beta_n$ are the coefficients associated with each predictor $x_1, x_2, \cdots x_n$ respectively

intercept

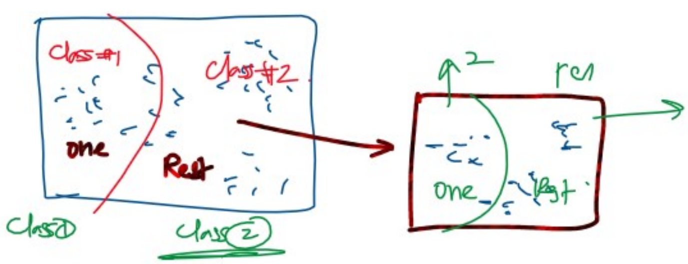
$$\beta_1 x_1 \rightarrow +\beta_1 x_1 \Rightarrow \log(\text{odds}) \uparrow\uparrow\uparrow$$

or

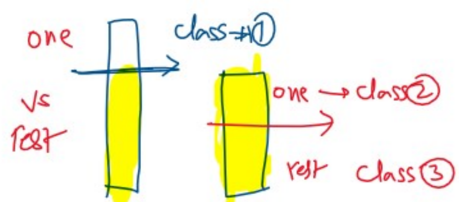
$$\beta_1 x_1 \rightarrow -\beta_1 x_1 \Rightarrow \log(\text{odds}) \downarrow\downarrow\downarrow$$

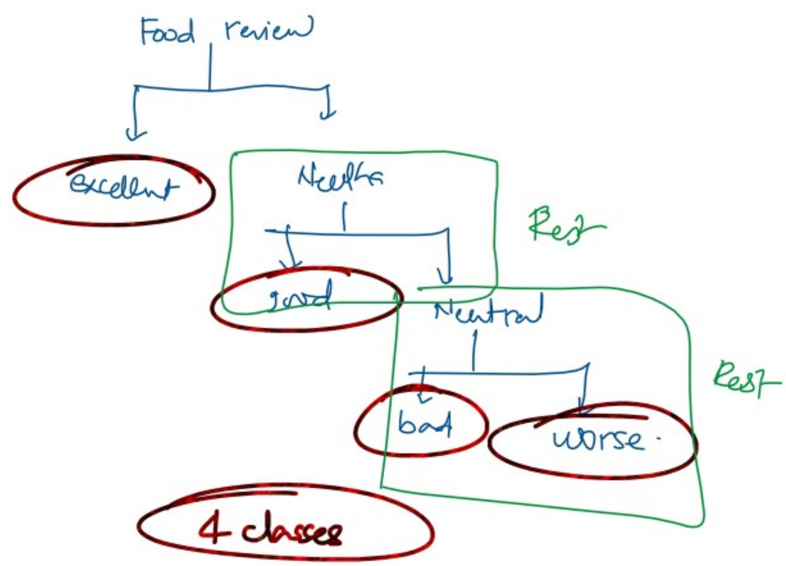
Logistic Regression for multiclass problems

one vs Rest (OVR)



4 class.





What is the purpose of intercept in logistic?

– all the predictor variables $(x_1, x_2 \cdots x_n)$ are set to zero, intercept (β_0) provides a

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

- to get baseline log of odds

log (odds) → 0.7 ← β_0

0.8 ← $\boxed{\beta_1 x_1}$

Assumptions of Logistic Regression

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

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

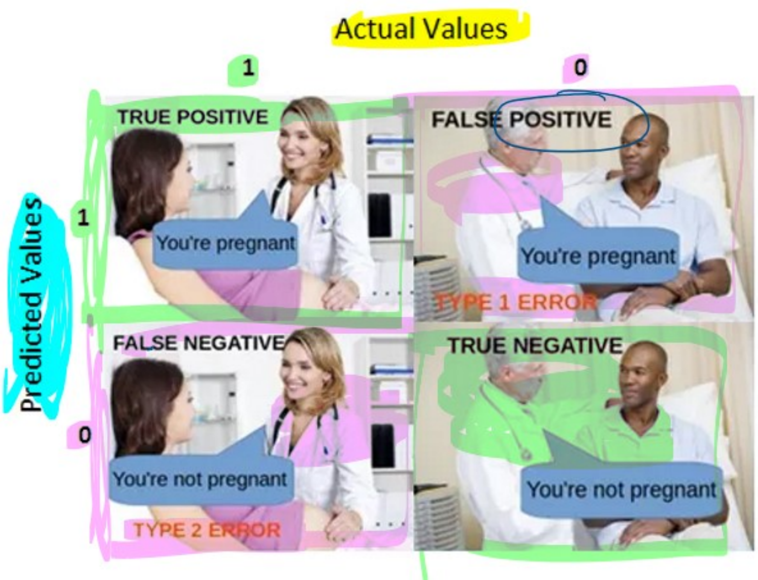
where output can be two or more classes.

Binary class

Actual Values

		Positive class (1)	Negative (0) class
Predicted values	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



Positive → Pregnant

Negative → Not Pregnant