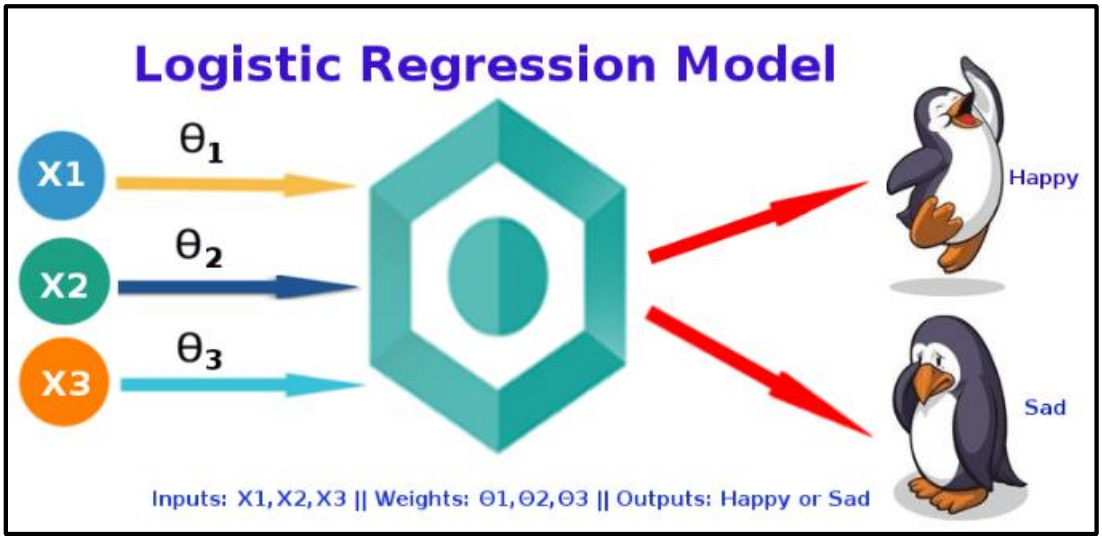
LOGISTIC REGRESSION NOTES

1) https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc

2) https://www.youtube.com/watch?v=-la3q9d7AKQ



Logistic Regression was used in the biological sciences in early twentieth century. It was then used in many social science applications. Logistic Regression is used when the dependent variable(target) is categorical.

For example,

- To predict whether an email is spam (1) or (0)
- Whether the tumor is malignant (1) or not (0)

Consider a scenario where we need to classify whether an email is spam or not.

If we use linear regression for this problem, there is a need for setting up a threshold based on which classification can be done.

Say if the actual class is malignant, predicted continuous value 0.4 and the threshold value is 0.5, the data point will be classified as not malignant which can lead to serious consequence in real time.

From this example, it can be inferred that linear regression is not suitable for classification problem. Linear regression is unbounded, and this brings logistic regression into picture.

Their value strictly ranges from 0 to 1.

Simple Logistic Regression

Model

Output = 0 or 1

Hypothesis => Z = WX + B

hΘ(x) = sigmoid (Z)

Sigmoid Function

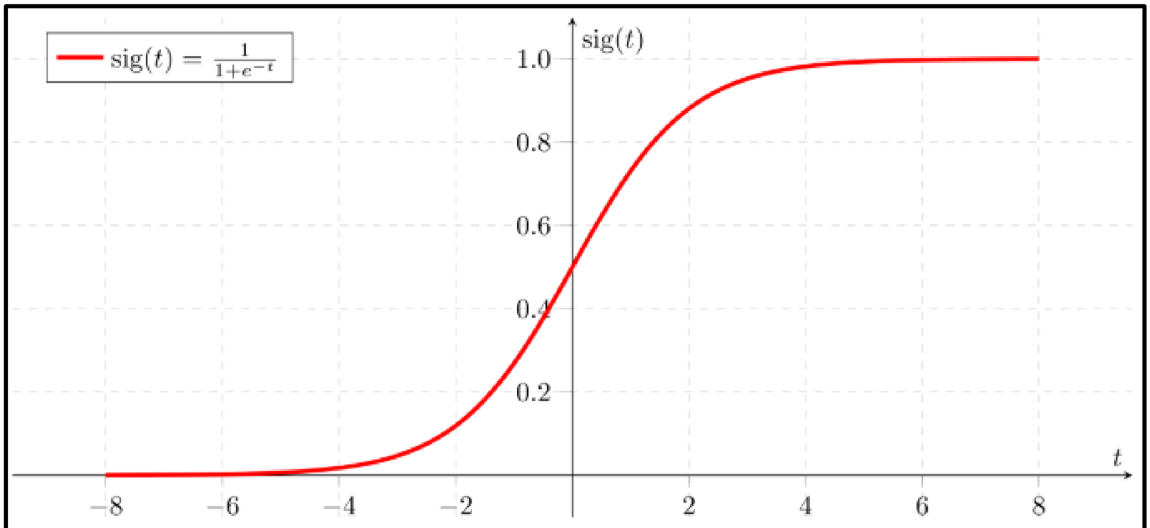


Figure 2: Sigmoid Activation Function

If ‘Z’ goes to infinity, Y(predicted) will become 1 and if ‘Z’ goes to negative infinity, Y(predicted) will become 0.

Analysis of the hypothesis

The output from the hypothesis is the estimated probability. This is used to infer how confident can predicted value be actual value when given an input X. Consider the below example,

X = [x0 x1] = [1 IP-Address]

Based on the x1 value, let’s say we obtained the estimated probability to be 0.8. This tells that there is 80% chance that an email will be spam.

Mathematically this can be written as,

$h_{\Theta}(x) = P(Y=1|X; theta)$

Probability that Y=1 given X which is parameterized by ‘theta’.

$P(Y=1|X; theta) + P(Y=0|X; theta) = 1$

$P(Y=0|X; theta) = 1 - P(Y=1|X; theta)$

Figure 3: Mathematical Representation

This justifies the name ‘logistic regression’. Data is fit into linear regression model, which then be acted upon by a logistic function predicting the target categorical dependent variable.

Types of Logistic Regression

1. **Binary Logistic Regression**: The categorical response has only two 2 possible outcomes. Example: Spam or Not

2. **Multinomial Logistic Regression:** Three or more categories without ordering. Example: Predicting which food is preferred more (Veg, Non-Veg, Vegan)

3. **Ordinal Logistic Regression**: Three or more categories with ordering. Example: Movie rating from 1 to 5.

Decision Boundary

- To predict which class a data belongs, a ***threshold can be set***.
- Based upon this threshold, **the obtained estimated probability is classified into classes**.
- Say, if predicted_value ≥ 0.5, then classify email as spam else as not spam.
- Decision boundary can be **linear or non-linear**.
- **Polynomial order can be increased to get complex decision boundary**.

Logistic Regression Cost Function

$$\text{Cost}(h_\Theta(x), Y(\text{actual})) = \begin{cases} -\log(h_\Theta(x)) & \text{if } y=1 \\ -\log(1-h_\Theta(x)) & \text{if } y=0 \end{cases}$$

Figure 4: Cost Function of Logistic Regression

Why cost function which has been used for linear cannot be used for logistic?

- Linear regression uses mean squared error as its cost function.
- If this is used for logistic regression, then it will be a non-convex function of parameters (theta).
- Gradient descent will converge into global minimum only if the function is convex.

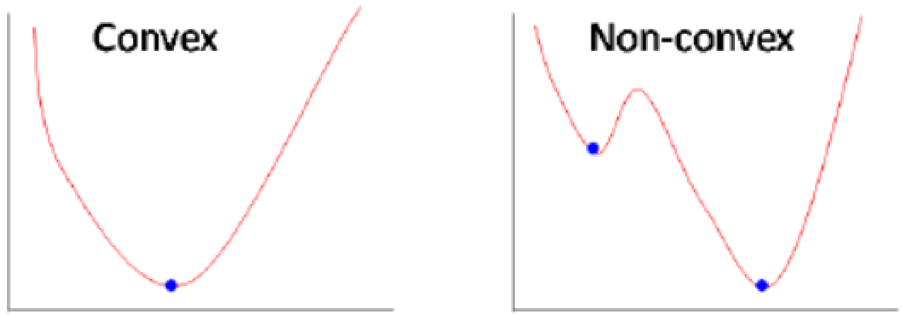
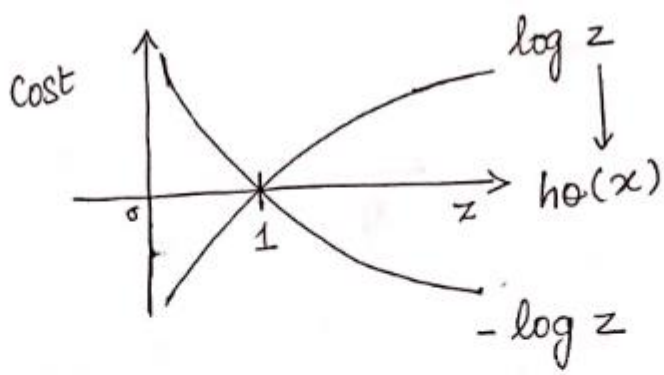


Figure 5: Convex and non-convex cost function

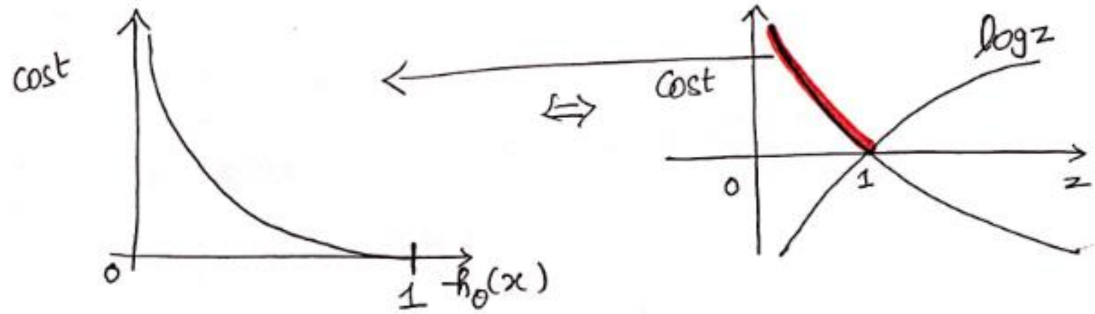
Cost function explanation



$$\mathrm{Cost}(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y = 1 \\ -\log(1 - h_\theta(x)) & \text{if } y = 0 \end{cases}$$

If $y = 1$,

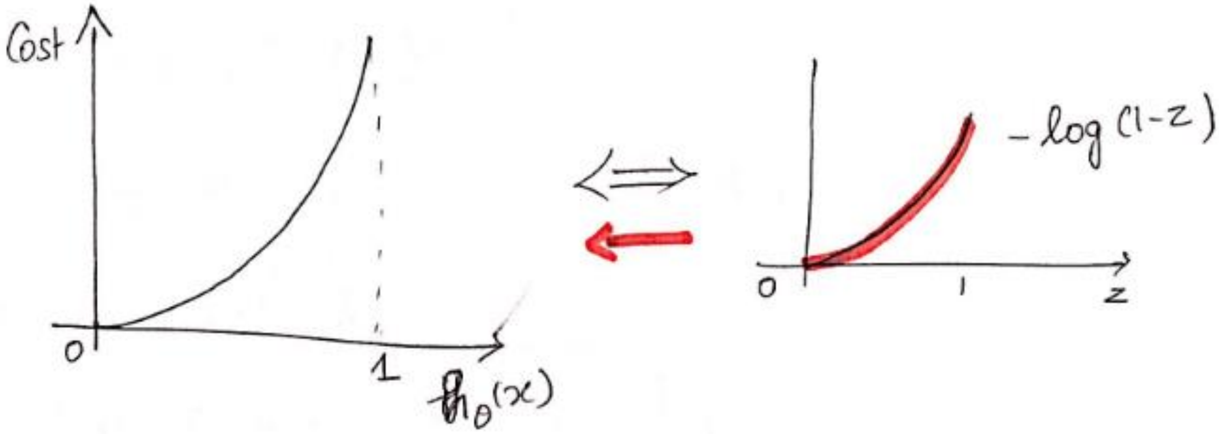
$$\text{Cost}(h_\theta(x), y) = -\log(h_\theta(x))$$



If, $\text{Cost} = 0 \Rightarrow y = 1 \Rightarrow h_\theta(x) = 1$

$\text{Cost} = \text{infinity for } h_\theta(x) = 0$

If $h_\theta(x) = 0$, it is similar to predicting $P(y=1 \mid x; \theta) = 0$



If $\text{cost} = 0 \Rightarrow h_\theta(x) = 0 \Rightarrow y = 0$

$\text{Cost} = \infty \Rightarrow h_\theta(x) = 1$

If $h_\theta(x) = 1$, it is similar to predicting

$P(y=0 \mid x; \theta) = 0$

Figure 7: Cost Function part 2

Simplified cost function

Cost($h_\Theta(x)$, y) = -y log($h_\Theta(x)$) – (1-y) log (1- $h_\Theta(x)$)

If y = 1, (1-y) term will become zero, therefore – log ($h_\Theta(x)$) alone will be present

If y = 0, (y) term will become zero, therefore – log (1- $h_\Theta(x)$) alone will be present

Figure 8: Simplified Cost Function

Why this cost function?

Let us consider,

* $\hat{y} = P(y=1|x)$

$\hat{y}$ is the probability that $y = 1$, given x

* $1-\hat{y} = P(y=0|x)$

* $P(y|x) = \hat{y}^{y} \cdot (1-\hat{y})^{(1-y)}$

If $y = 1 \Rightarrow P(y|x) = \hat{y}$

Figure 9: Maximum Likelihood Explanation part-1

$$\Rightarrow \quad \log\left(\hat{y}^{\,y} \cdot (1-\hat{y})^{(1-y)}\right)$$

$$\Rightarrow \quad y \log \hat{y} + (1-y) \log (1-\hat{y})$$

$$\Rightarrow \quad -L(\hat{y}, \hat{y})$$

$$\boxed{\log P(y|x) = -L(\hat{y}, y)}$$

Figure 10: Maximum Likelihood Explanation part-2

This negative function is because when we train, we need to maximize the probability by minimizing loss function. Decreasing the cost will increase the maximum likelihood assuming that samples are drawn from an identically independent distribution.

Deriving the formula for Gradient Descent Algorithm

Gradient

$$\boxed{z = w_1 x_1 + w_2 x_2 + b} \longrightarrow \boxed{\hat{y} = a = \sigma(z)} \longrightarrow \boxed{L(\hat{y}, y)}$$

$$\Leftrightarrow \boxed{a = \hat{y}}$$

$$\boxed{w_1} \Rightarrow \frac{\partial(L)}{\partial w_1} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z} \cdot \frac{\partial(z)}{\partial w_1}$$

$$\frac{\partial L}{\partial a} = \frac{\partial}{\partial a}\left(-y \log a - (1-y)\log(1-a)\right)$$

$$= -y\left(\frac{1}{a}\right) - (-1)\frac{(1-y)}{(1-a)}$$

$$\boxed{\frac{\partial L}{\partial a} = \left(\frac{-y}{a}\right) + \left(\frac{1-y}{1-a}\right)}$$

$$\boxed{\frac{\partial a}{\partial z} = a(1-a)}$$

$$\boxed{\frac{\partial z}{\partial w_1} = x_1}$$

$$\frac{\partial L}{\partial w_1} = \left(\left(\frac{-y}{a} + \frac{(1-y)}{1-a} \right) \cdot (a)(1-a) \right) \cdot x_1$$

$$= (a-y) \cdot x_1$$

Update for w_1,

$$\frac{\partial L}{\partial w_1} = (a-y) \cdot x_1$$

Here, $(a-y) = \frac{\partial L}{\partial z}$

$$\boxed{w_1 = w_1 - \alpha \frac{\partial L}{\partial w_1}}$$

Similarly, for all parameters

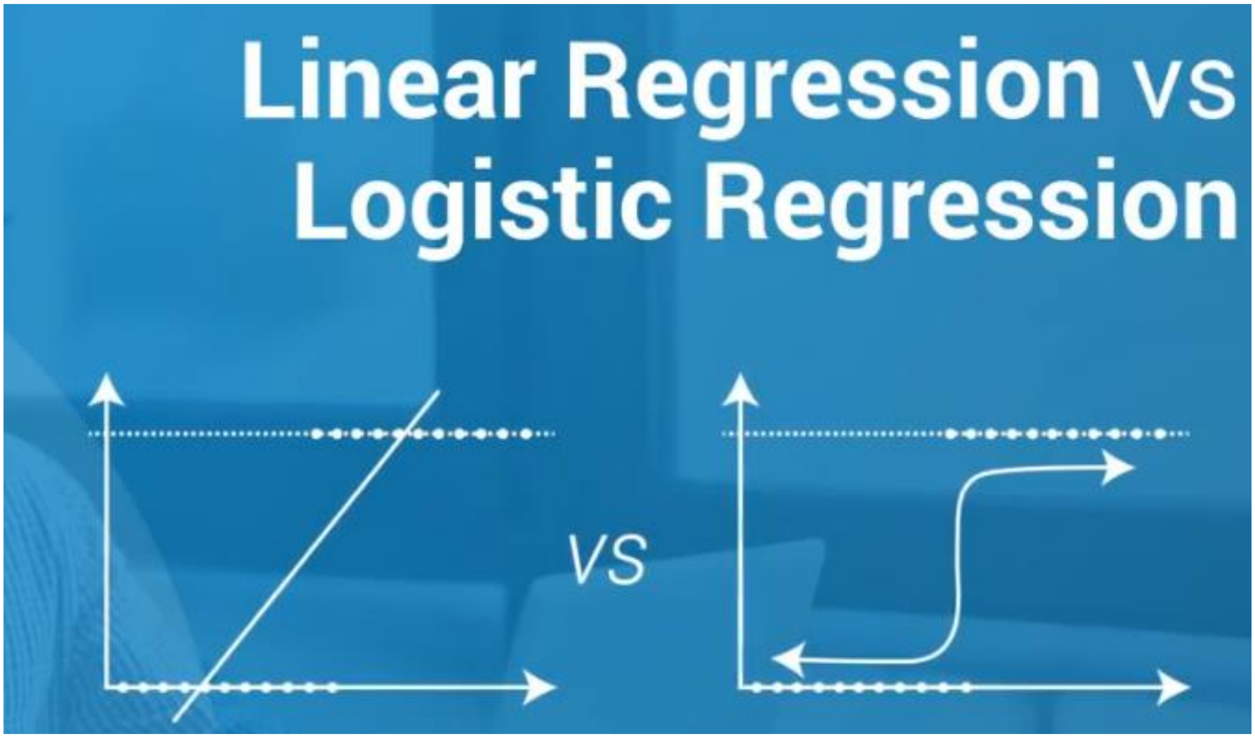
$$\boxed{w_i = w_i - \alpha \frac{\partial L}{\partial w_i}}$$

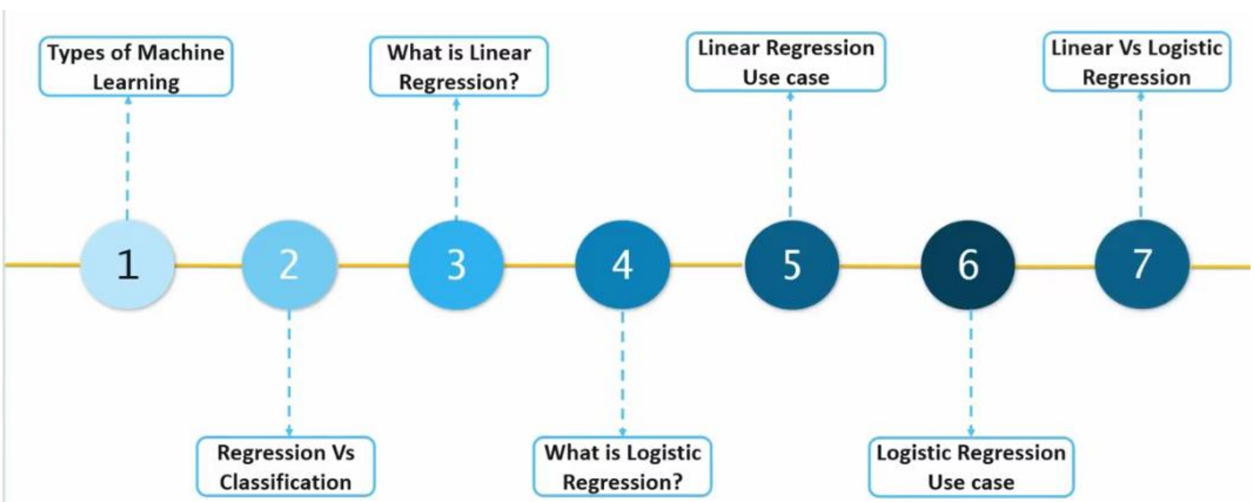
$i = 1, 2, \ldots, m$

m = no of parameters

$$\boxed{b = b - \alpha \frac{\partial L}{\partial b}}$$

Where, $\frac{\partial L}{\partial b} = (a-y)$


Linear Regression vs
Logistic Regression
VS


Types of Machine Learning
1
2
Regression Vs Classification
What is Linear Regression?
3
4
What is Logistic Regression?
Linear Regression Use case
5
6
Logistic Regression Use case
Linear Vs Logistic Regression
7

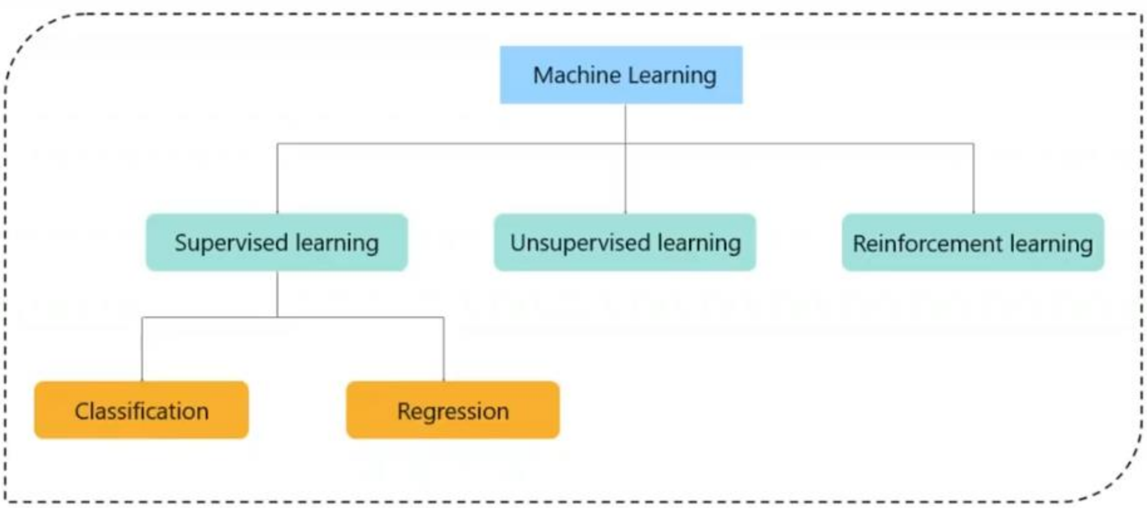
Types Of Machine Learning

Supervised Learning

Unsupervised Learning

Reinforcement Learning

Regression And Classification

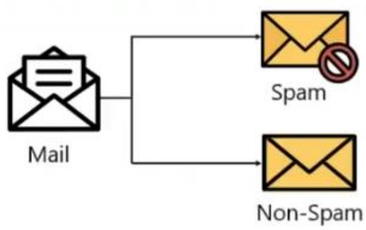


Regression Vs Classification

Classification

Classification is the task of predicting a discrete class label

- In a classification problem data is classified into one of two or more classes
- A classification problem with two classes is called binary, more than two classes is called a multi-class classification



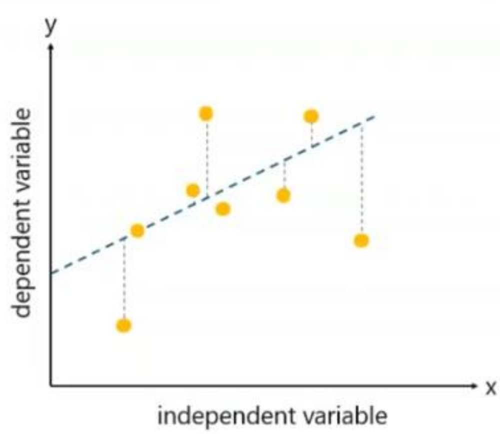
Regression

Regression is the task of predicting a continuous quantity

- A regression problem requires the prediction of a quantity
- A regression problem with multiple input variables is called a multivariate regression problem

What Is Linear Regression?

Linear Regression is a method to predict dependent variable (Y) based on values of independent variables (X). It can be used for the cases where we want to predict some continuous quantity.



- *Dependent variable (Y):*

The response variable who's value needs to be predicted.

- *Independent variable (X):*

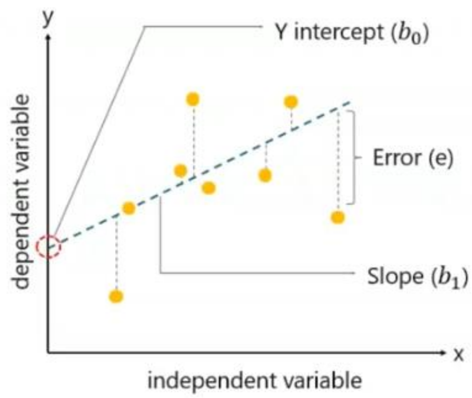
The predictor variable used to predict the response variable.

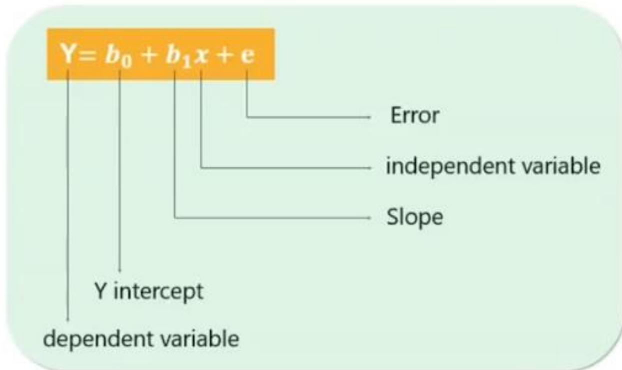
The following equation is used to represent a linear regression model:

$$Y = b_0 + b_1 x + e$$

What Is Linear Regression?

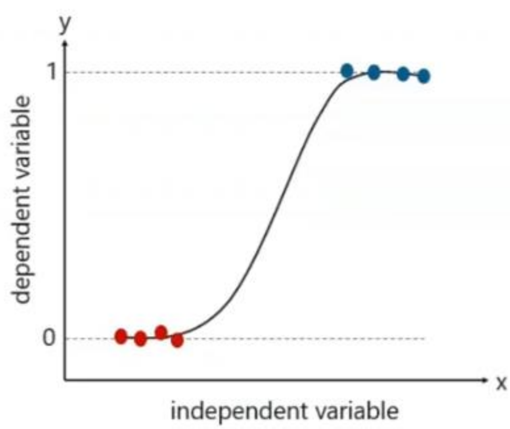
Linear Regression is a method to predict dependent variable (Y) based on values of independent variables (X). It can be used for the cases where we want to predict some continuous quantity.





What Is Logistic Regression?

Logistic Regression is a method used to predict a dependent variable, given a set of independent variables, such that the dependent variable is categorical.



- *Dependent variable (Y):*
The response binary variable holding values like 0 or 1, Yes or No, A, B or C

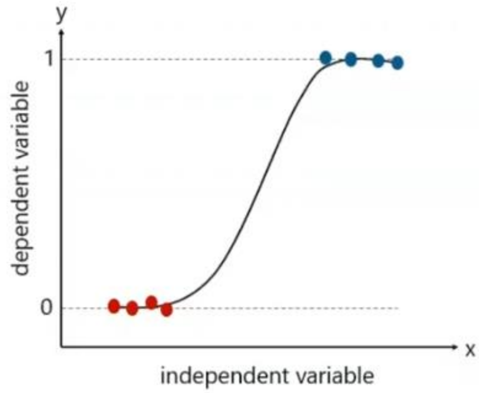
- *Independent variable (X):*
The predictor variable used to predict the response variable.

The following equation is used to represent a linear regression model:

$$\log\left(\frac{Y}{1-Y}\right) = C + B1X1 + B2X2 + \ldots.$$

What Is Logistic Regression?

Logistic Regression is a method used to predict a dependent variable, given a set of independent variables, such that the dependent variable is categorical.



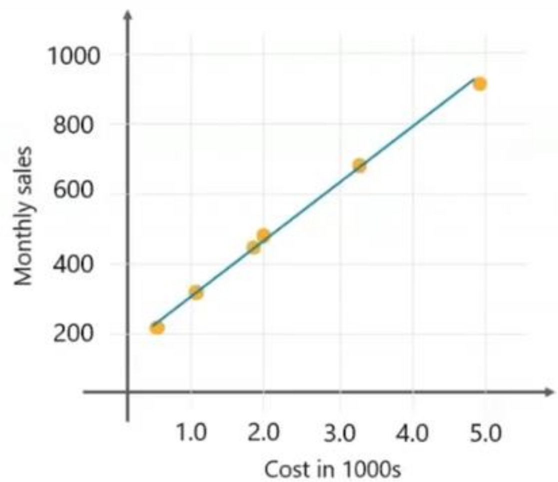
$$\log\left(\frac{Y}{1-Y}\right) = C + B1X1 + B2X2 + ...$$

- Y is the probability of an event to happen which you are trying to predict
- x1, x2 are the independent variables which determine the occurrence of an event i.e. Y
- C is the constant term which will be the probability of the event happening when no other factors are considered

Linear Regression Use Case

To forecast monthly sales by studying the relationship between the monthly e-commerce sales and the online advertising costs.

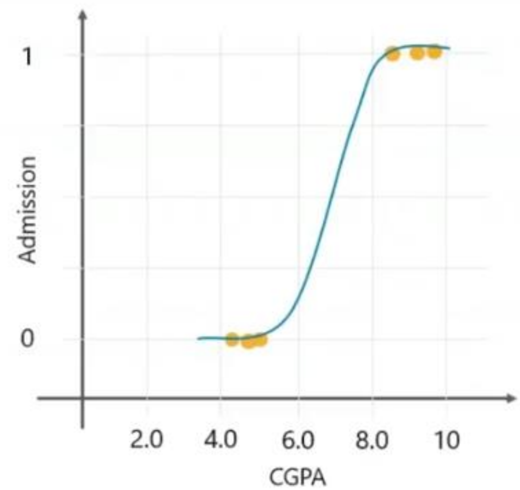
Monthly sales	Advertising cost In 1000s
200	0.5
900	5
450	1.9
680	3.2
490	2.0
300	1.0



Logistic Regression Use Case

To predict if a student will get admitted to a school based on his CGPA.

Admission	CGPA
0	4.2
0	5.1
0	5.5
1	8.2
1	9.0
1	9.1



Linear Regression Vs Logistic Regression

	Linear Regression	Logistic Regression
1 Definition	To predict a continuous dependent variable based on values of independent variables	*To predict a categorical dependent variable based on values of independent variables*
2 Variable Type	Continuous dependent variable	Categorical dependent variable
3 Estimation method	Least square estimation	Maximum like-hood estimation
4 Equation	$Y = b_0 + b_1 x + e$	$\log\left(\frac{Y}{1-Y}\right) = C + B1X1 + B2X2 + \ldots$
5 Best fit line	Straight line	Curve
6 Relationship between DV & IV	Linear relationship between the dependent and independent variable	Linear relationship is not mandatory
7 Output	Predicted integer value	Predicted binary value (0 or 1)
8 Applications	Business domain, forecasting sales	Classification problems, cybersecurity, image processing