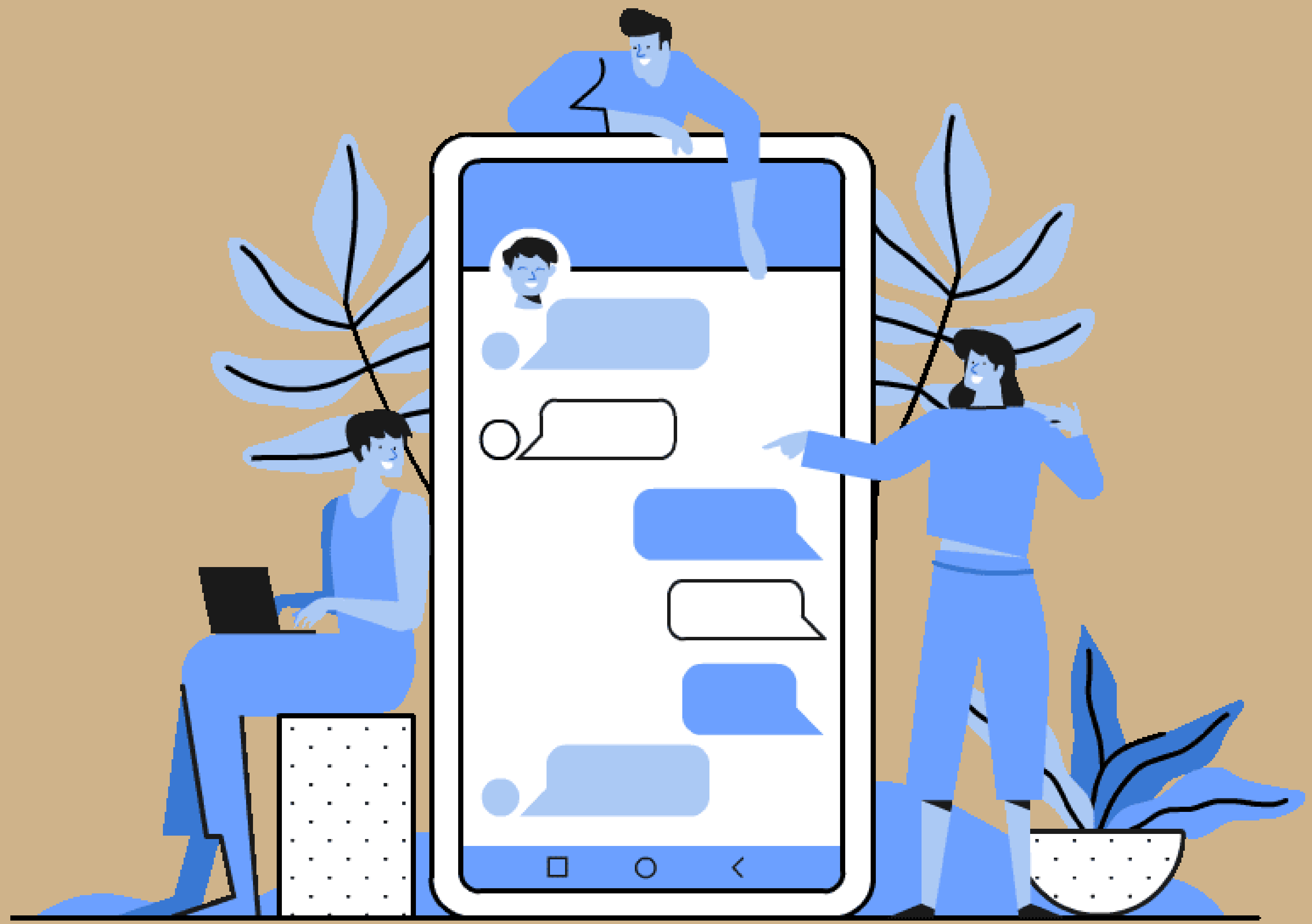
Logistic Regression

Tag - J

Let's get started!



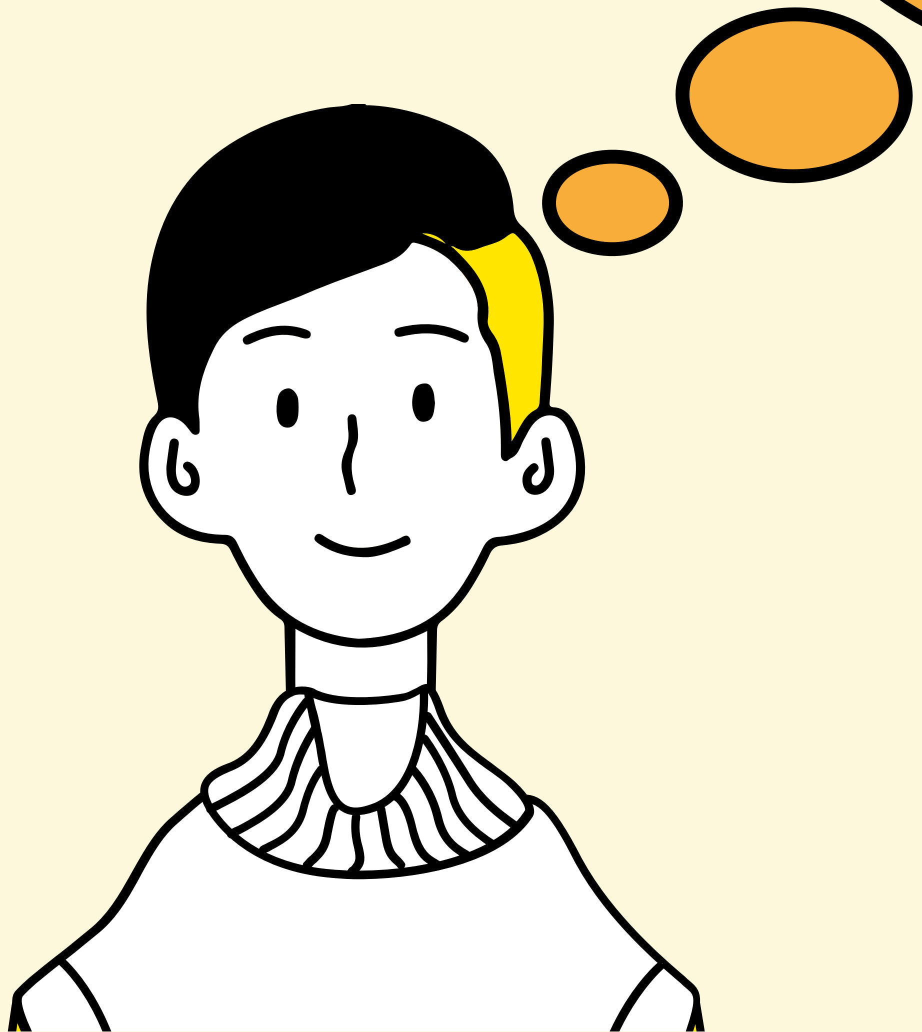
The factor that is being predicted is called the dependent variable. The factors that are used to predict the value of the dependent variable are called the independent variables.

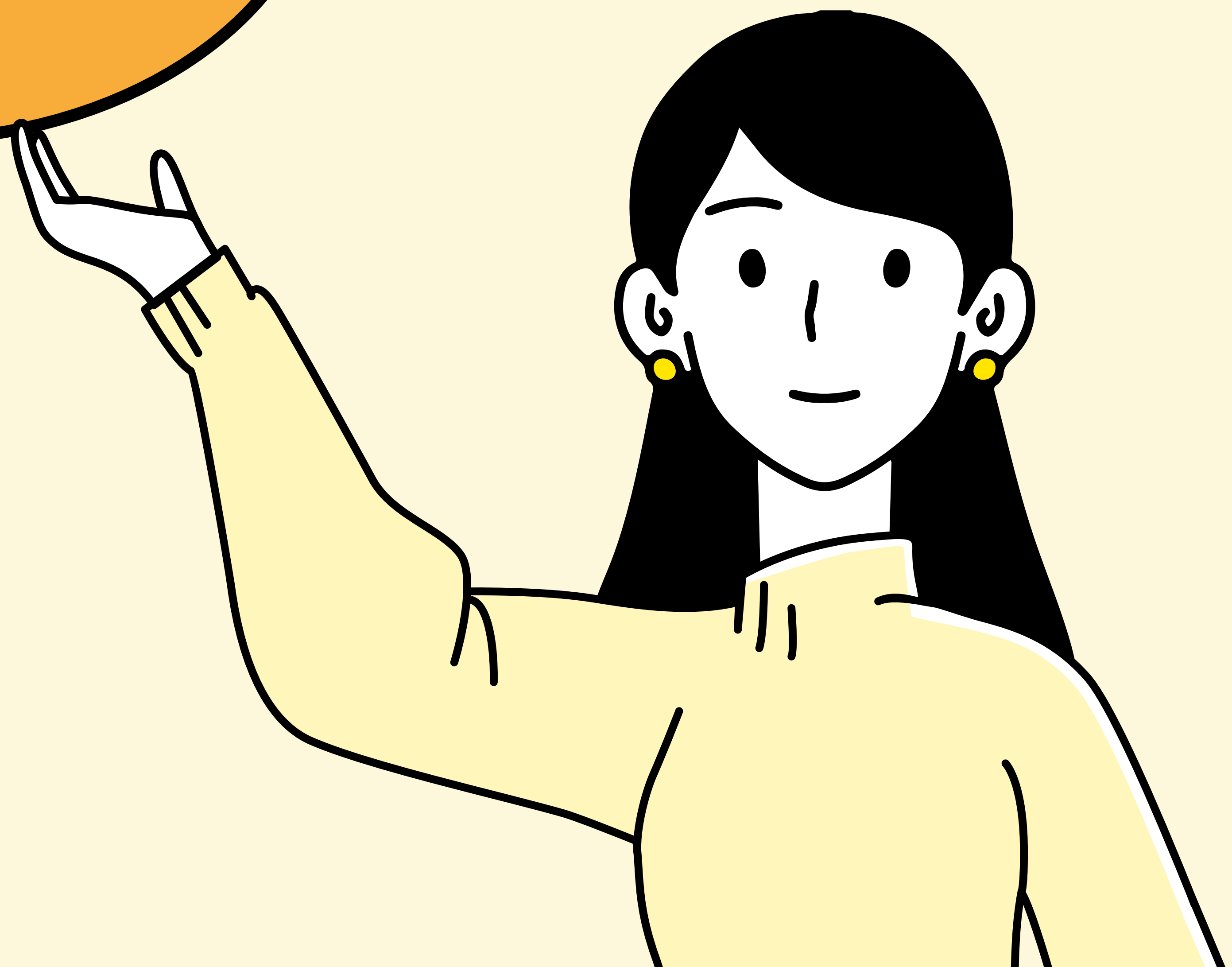
Linear regression models are used to show or predict the relationship between two variables or factors.

In linear regression, each observation consists of two values. One value is for the dependent variable and one value is for the independent variable.

In this simple model, a straight line approximates the relationship between the dependent variable and the independent variable.

When two or more independent variables are used in regression analysis, then the model is known as multiple regression.





What is Logistic Regression?

- Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable.
- The nature of target or dependent variable is dichotomous, which means there would be only two possible classes.

2

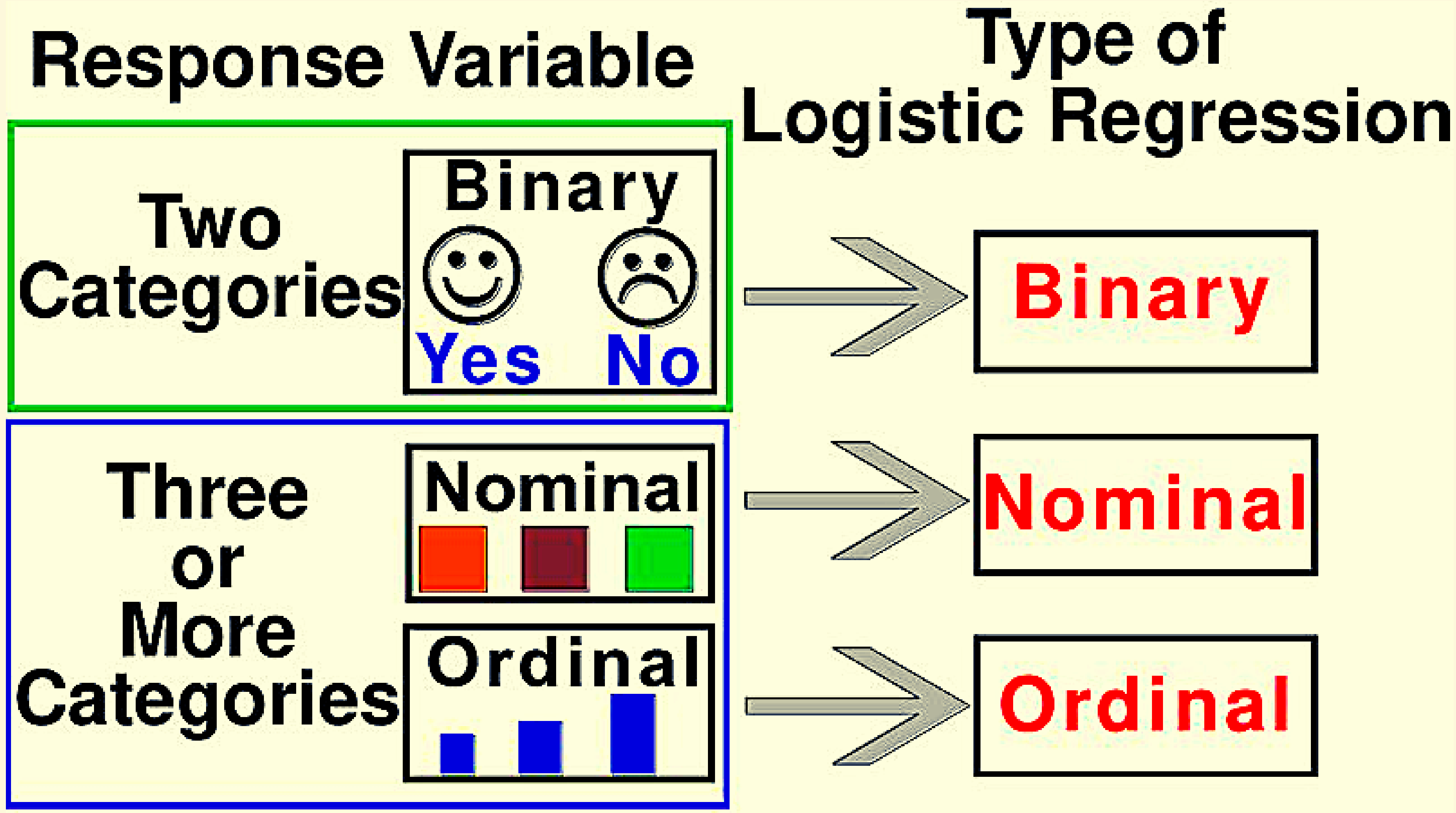
Mathematically, a logistic regression model predicts P(Y=1) as a function of X.

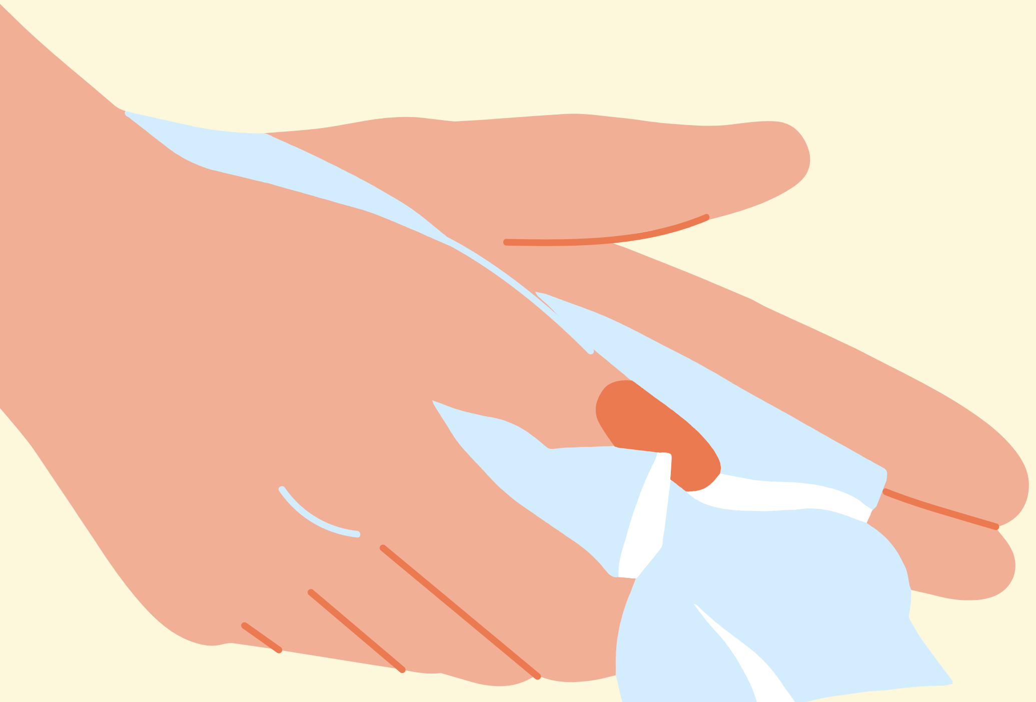
1

The dependent variable is binary in nature having data coded as either 1 (stands for success/yes) or 0 (stands for failure/no).

3

It is one of the simplest ML algorithms that can be used for various classification problems such as spam detection, Diabetes prediction, cancer detection etc.

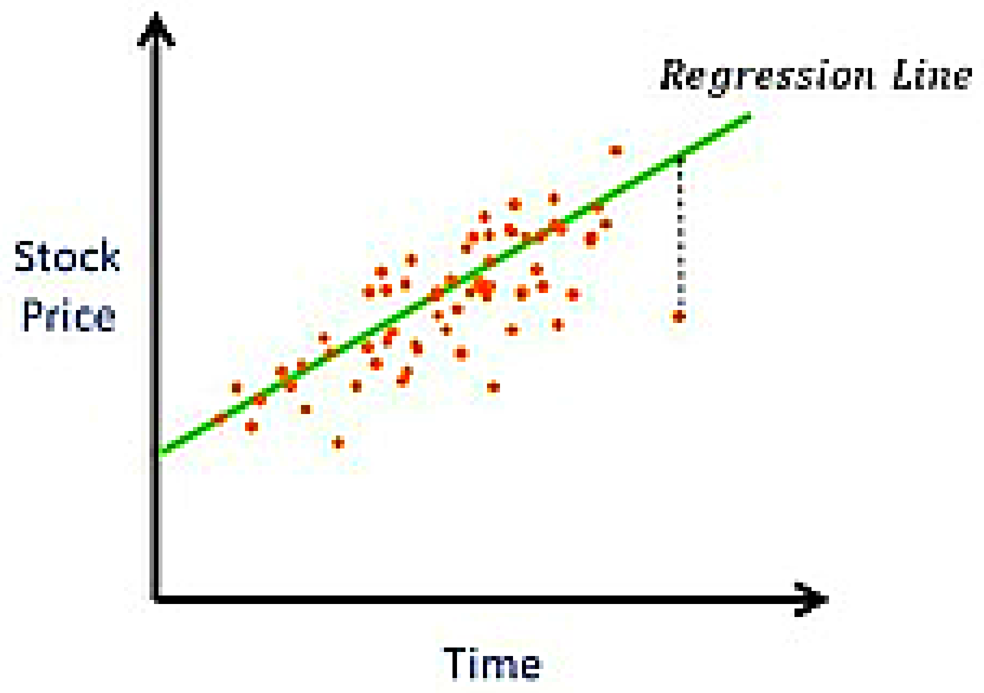

Response Variable
Type of Logistic Regression
Two Categories
Binary
Yes
No
Binary
Three or More Categories
Nominal
Ordinal
Nominal
Ordinal



Difference

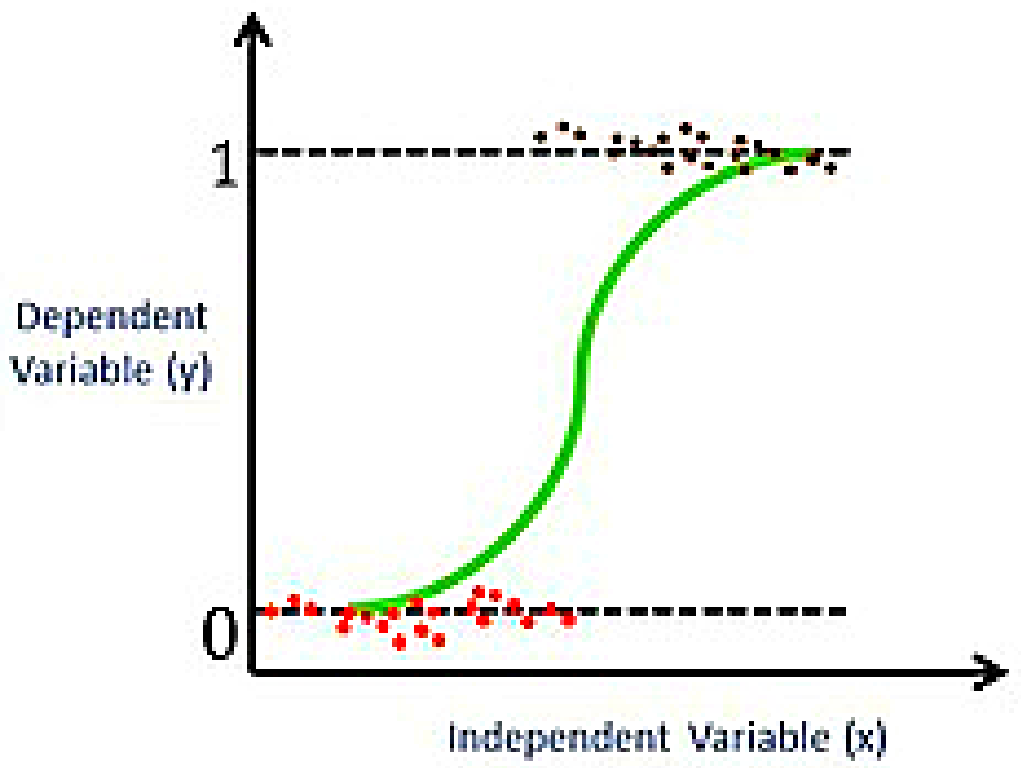
Linear Regression

- Aim is to predict continuous valued output.
- Output value can be any possible integer number.



Logistic Regression

- Aim is to predict the label for input data.
- Output is categorical (Binary) i.e. 0/1, True/False, etc.



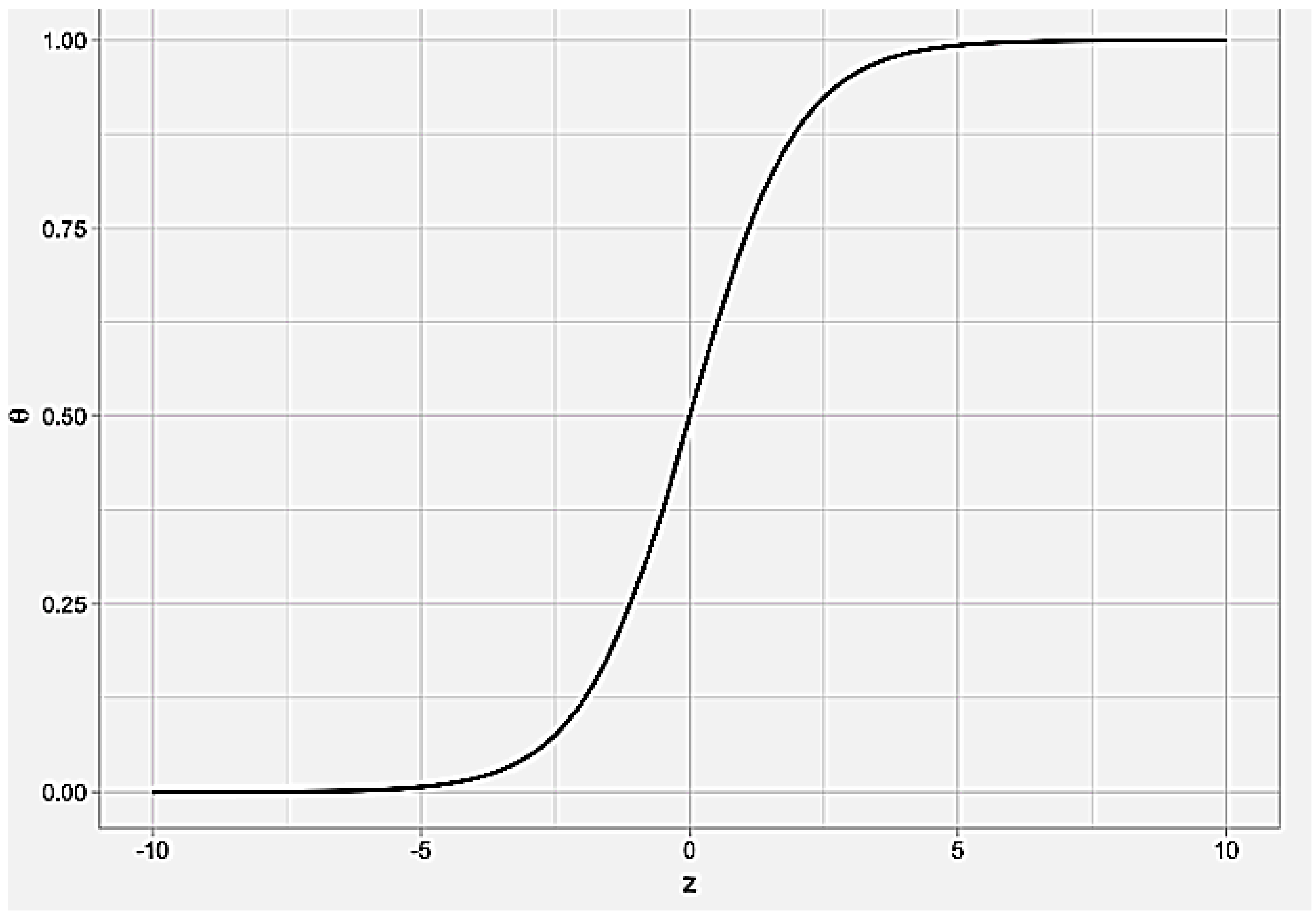
Sigmoid Function

- Let z be any continuous value whose domain is (-∞,∞).
- If you plug z into the sigmoid function like

$$\theta(z) = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}$$

- A nice property of the output is that it is always within 0 and 1.

Sigmoid Function



Here are some properties of θ(z):

When z = 0, θ =0.5

$$\theta(z) = \frac{1}{1+e^0} = \frac{1}{1+1} = .5$$

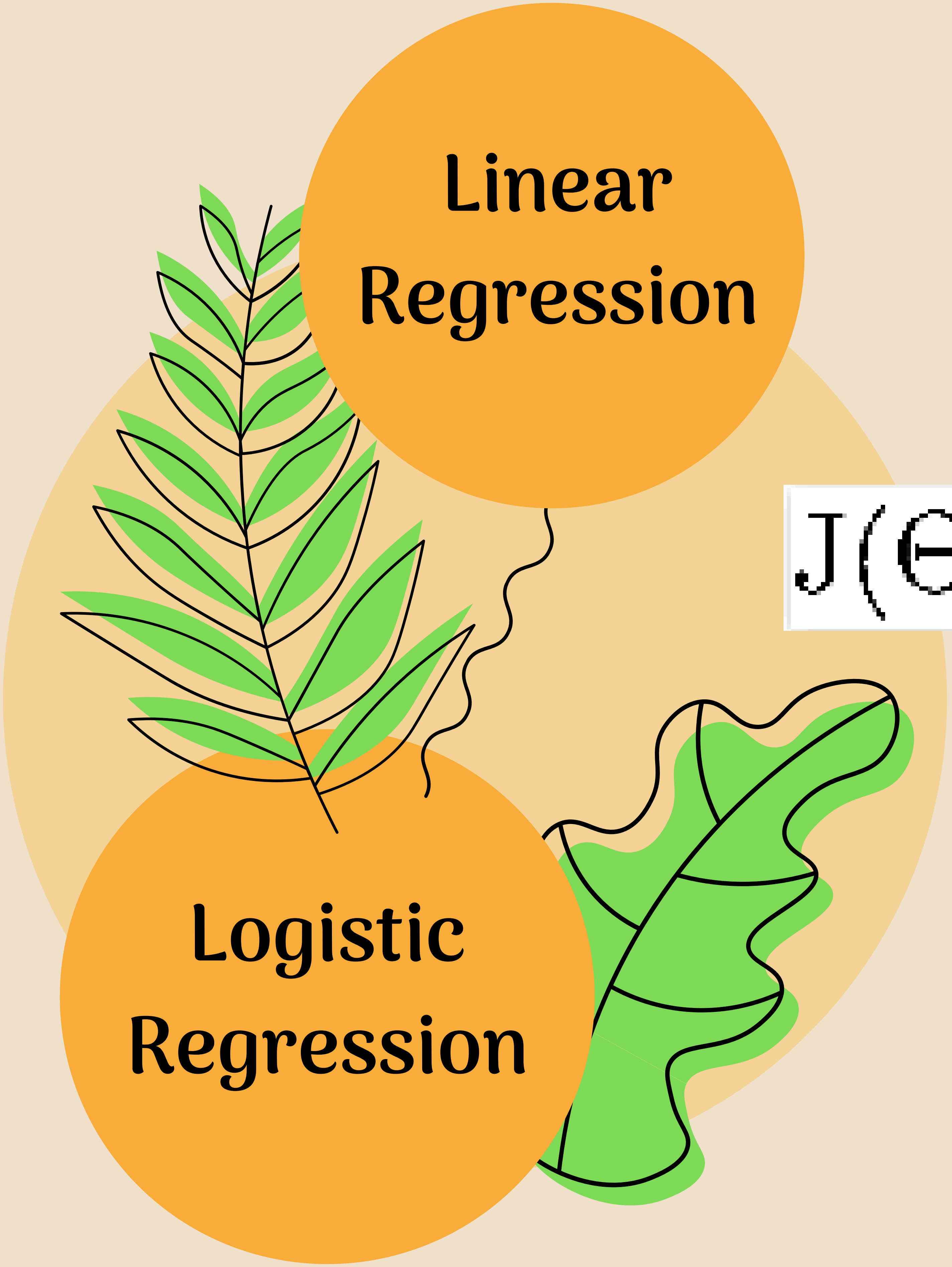
When z is very large, **θ** is approximately 1

$$\theta(z) \approx \frac{1}{1+0} = 1$$

3.When z is very small/negative, θ is approximately 0

$$\theta(z) \approx \frac{1}{1+\infty} = 0$$

We can use the sigmoid function to convert a continuous, unbounded output z to a decimal number $\theta \in (0,1)$, which is advantageous for representing probabilities.



Cost Function

$$J(\Theta) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{2}\left[h_\Theta(x^{(i)}) - y^{(i)}\right]^2$$

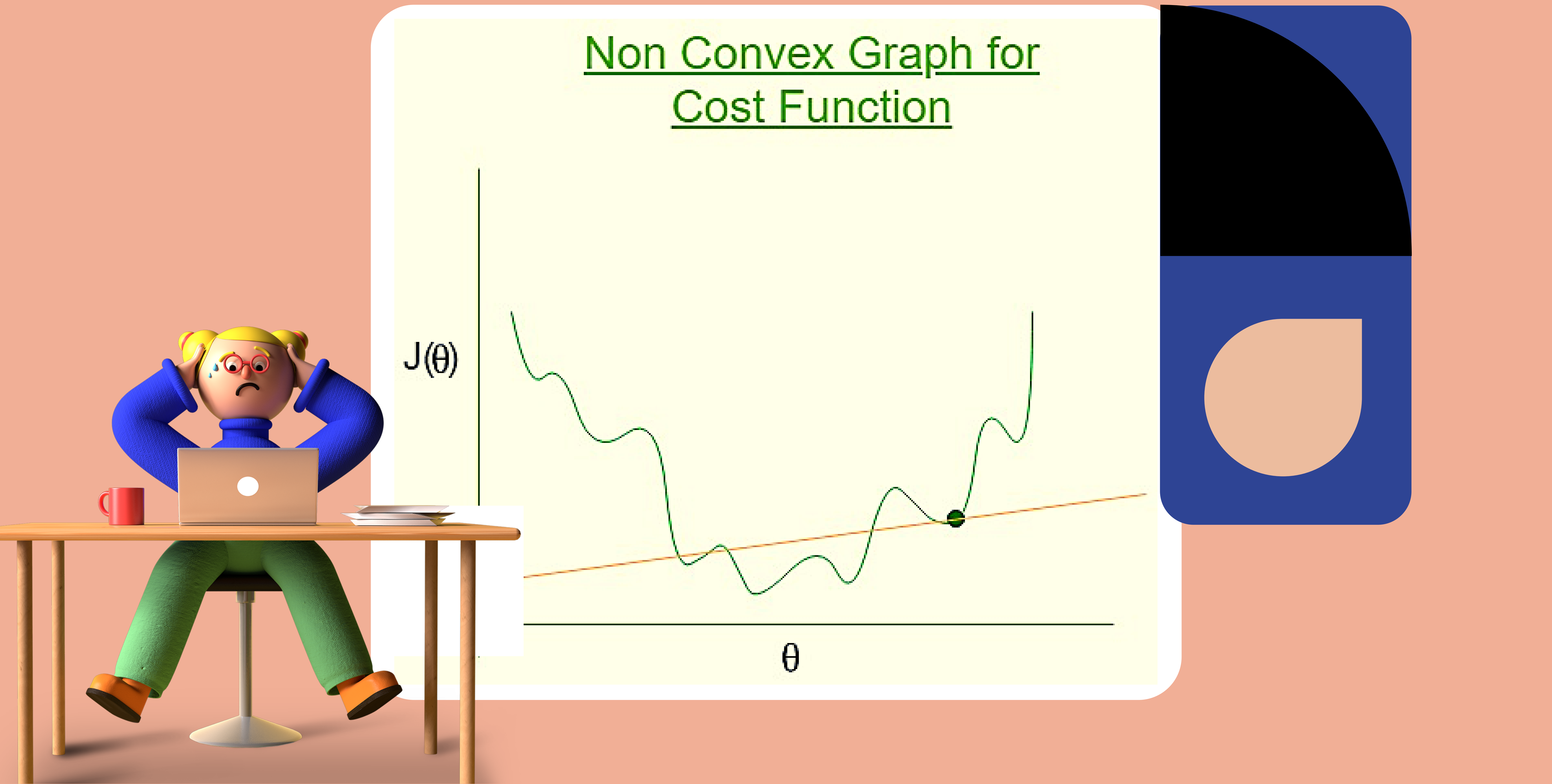
$$h_\Theta(x) = g(\Theta^T x)$$

Cost Function

$$\mathrm{Cost}(h_\Theta(x), y) = \begin{cases} 0 & if \quad h_\Theta(x) = y \\ \infty & if \quad y = 0 \quad and \quad h_\Theta(x) \rightarrow 1 \\ \infty & if \quad y = 1 \quad and \quad h_\Theta(x) \rightarrow 0 \end{cases}$$

$$J(\Theta) = \frac{-1}{m} \sum_{i=1}^{m} Cost(h_\Theta(x), y)$$

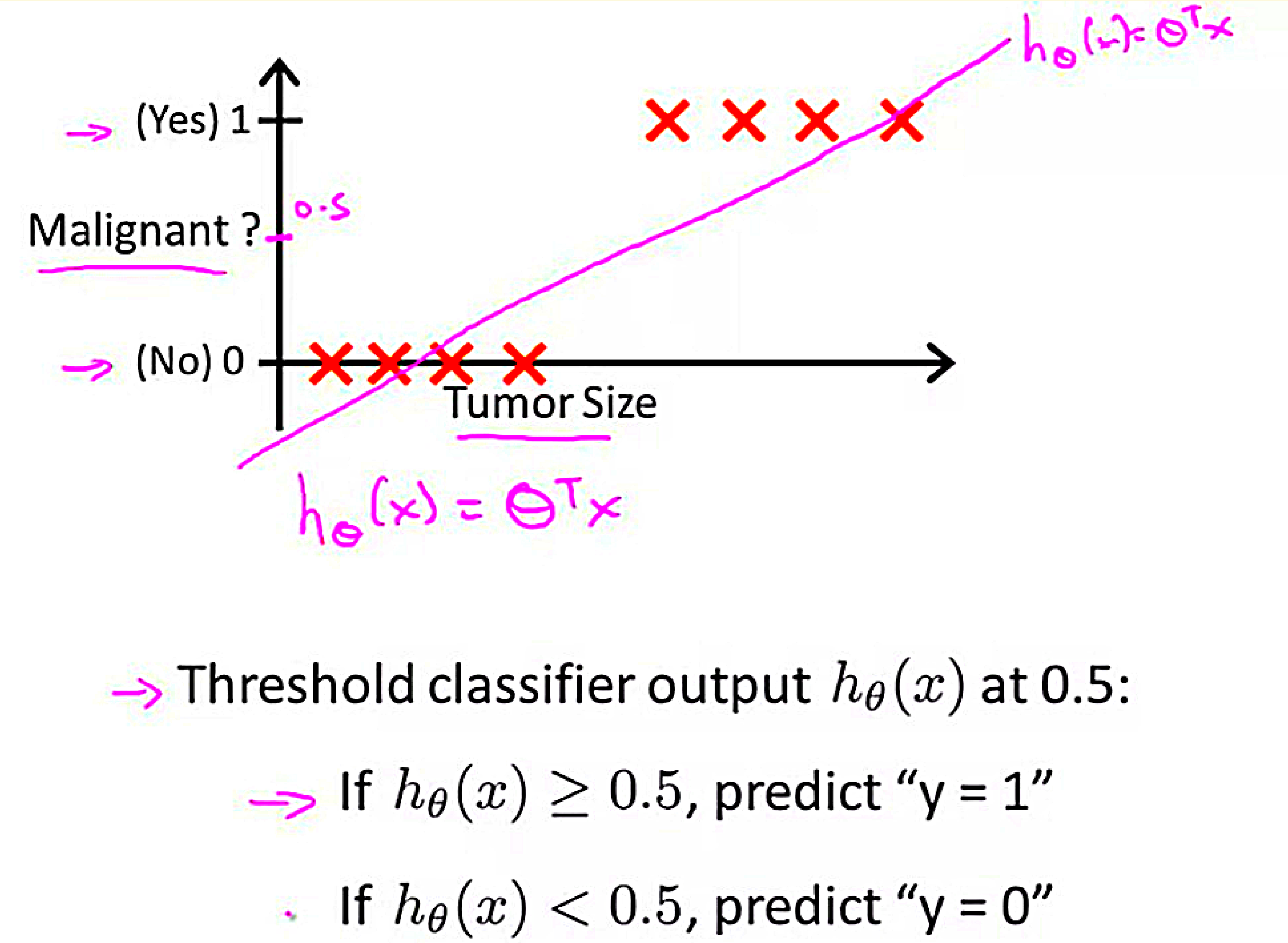
To fit parameter θ, J(θ) has to be minimized and for that Gradient Descent is required.


Non Convex Graph for
Cost Function
J(θ)
θ

Examples

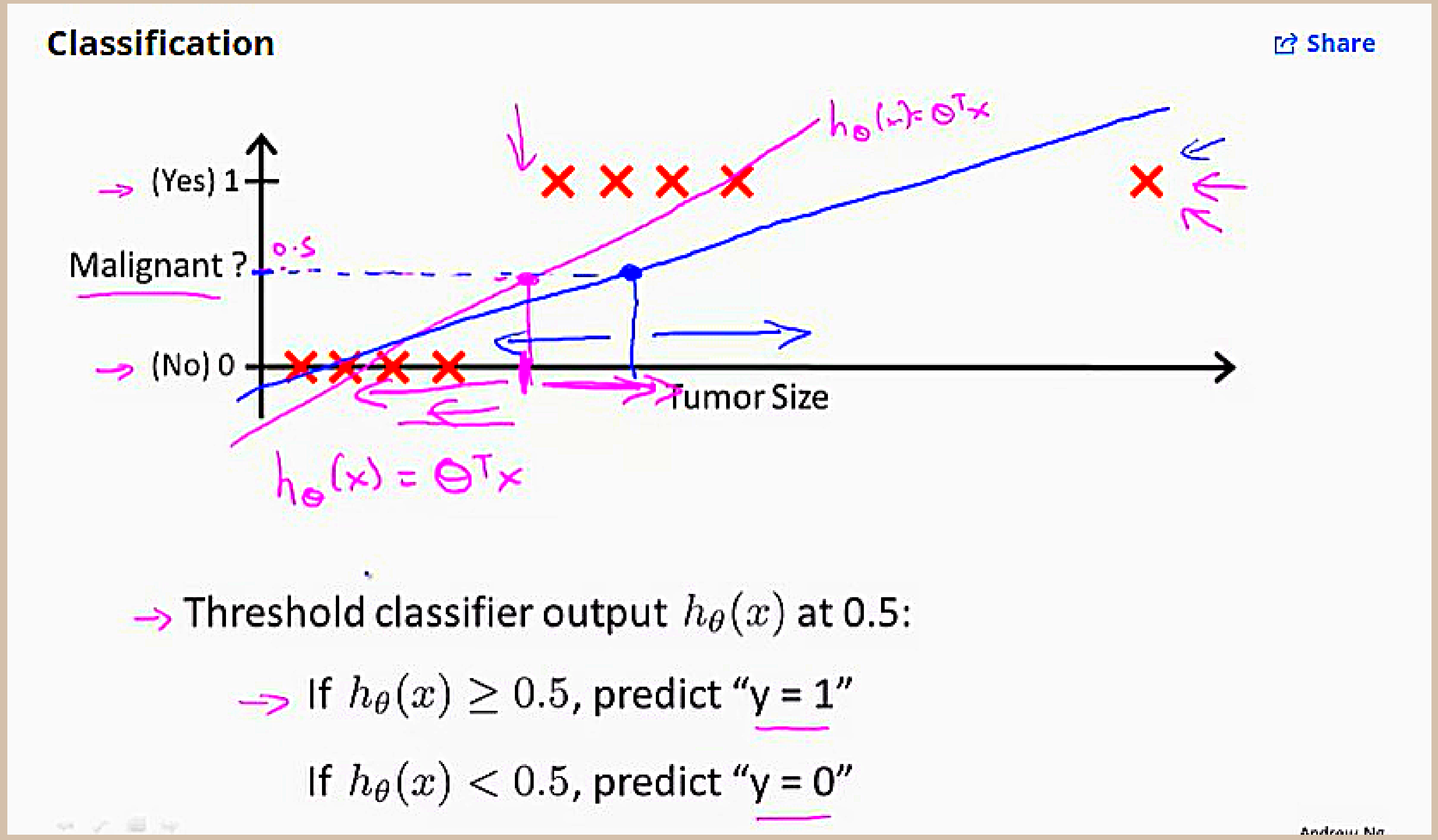
Let us consider an example of cancer prediction whether it is malignant or not.And label Malignant or not on y-axis and tumor size on x-axis.

<u>In linear regression:</u>



- h(x) is proportional to tumor size
- And h(x) depends on training set

- Because of this, we loose the accuracy in prediction whether it is malignant or not.
- And also h(x)>1 and h(x)<0 which is unbound
- So we need a function that makes 0<=h(x)<=1



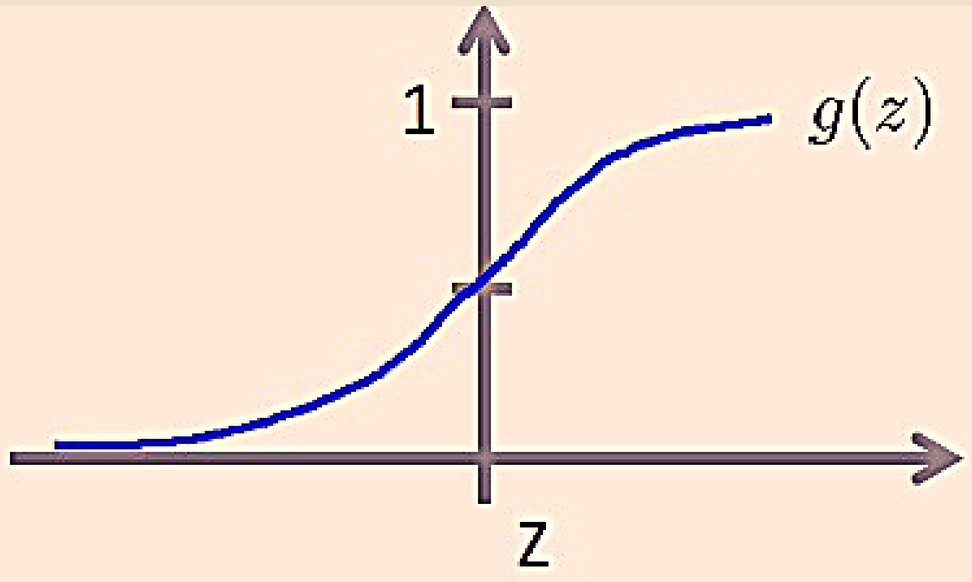
So we use Logistic regression for classification problems

In Logistic Regression:

Logistic regression

$$h_\theta(x) = g(\theta^T x)$$

$$g(z) = \frac{1}{1+e^{-z}}$$



Suppose predict "$y = 1$" if $h_\theta(x) \geq 0.5$

$\theta^T x \geq 0$

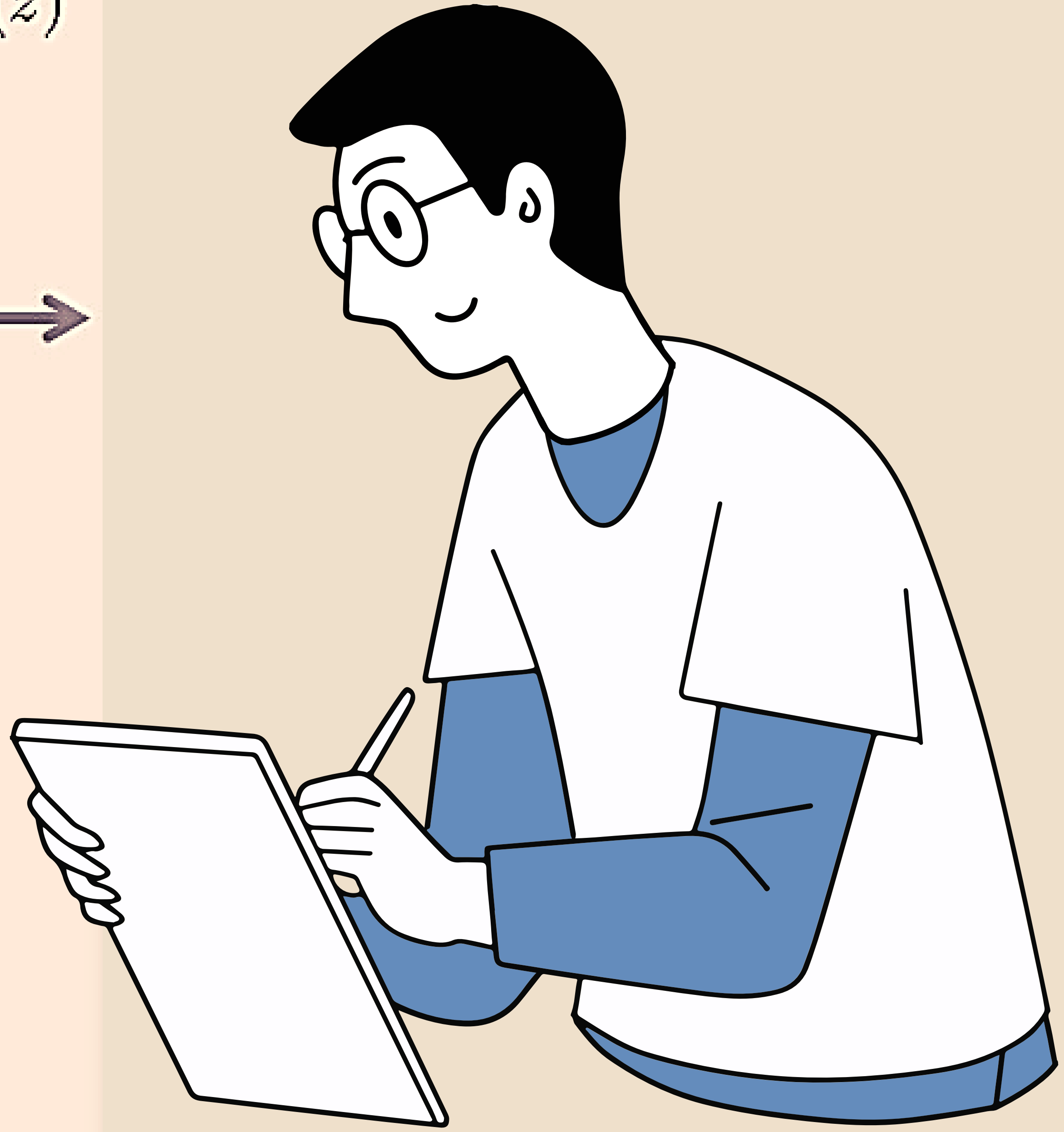
$g(z) \geq 0.5$
when $z \geq 0$

$h_\theta(x) = g(\theta^T x$

predict "$y = 0$" if $h_\theta(x) < 0.5$

$\theta^T x < 0$

$g(z) < 0.5$
when $z < 0$



Advantages

- Logistic regression is easier to implement, interpret, and very efficient to train
- It makes no assumptions about distributions of classes in feature space
- It can easily extend to multiple classes(multinomial regression) and a natural probabilistic view of class predictions.
- It not only provides a measure of how appropriate a predictor(coefficient size)is, but also its direction of association (positive or negative)..

Dis Advantage

- If the number of observations is lesser than the number of features, Logistic Regression should not be used, otherwise, it may lead to overfitting.
- It constructs linear boundaries.
- The major limitation of Logistic Regression is the assumption of linearity between the dependent variable and the independent variables.
- It can only be used to predict discrete functions. Hence, the dependent variable of Logistic Regression is bound to the discrete number set.


Thank you!