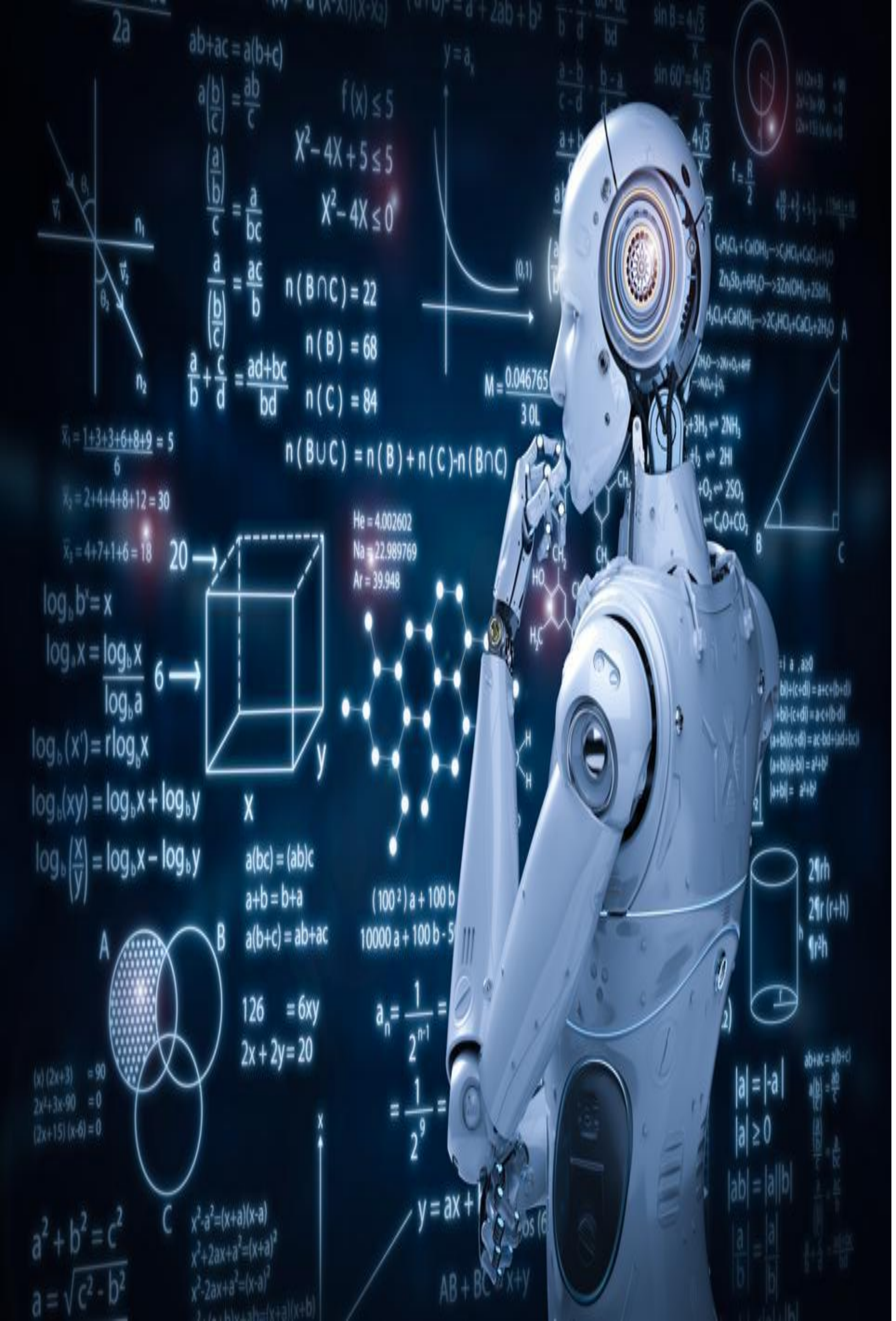


Practical Machine Learning

Day 6: Mar22 DBDA

Kiran Waghmare

Agenda

- Ridge, Lasso &ElasticNet
- Preprocessing Techniques
- Classification

Simple Linear Regression

$$y = b_0 + b_1 x_1$$

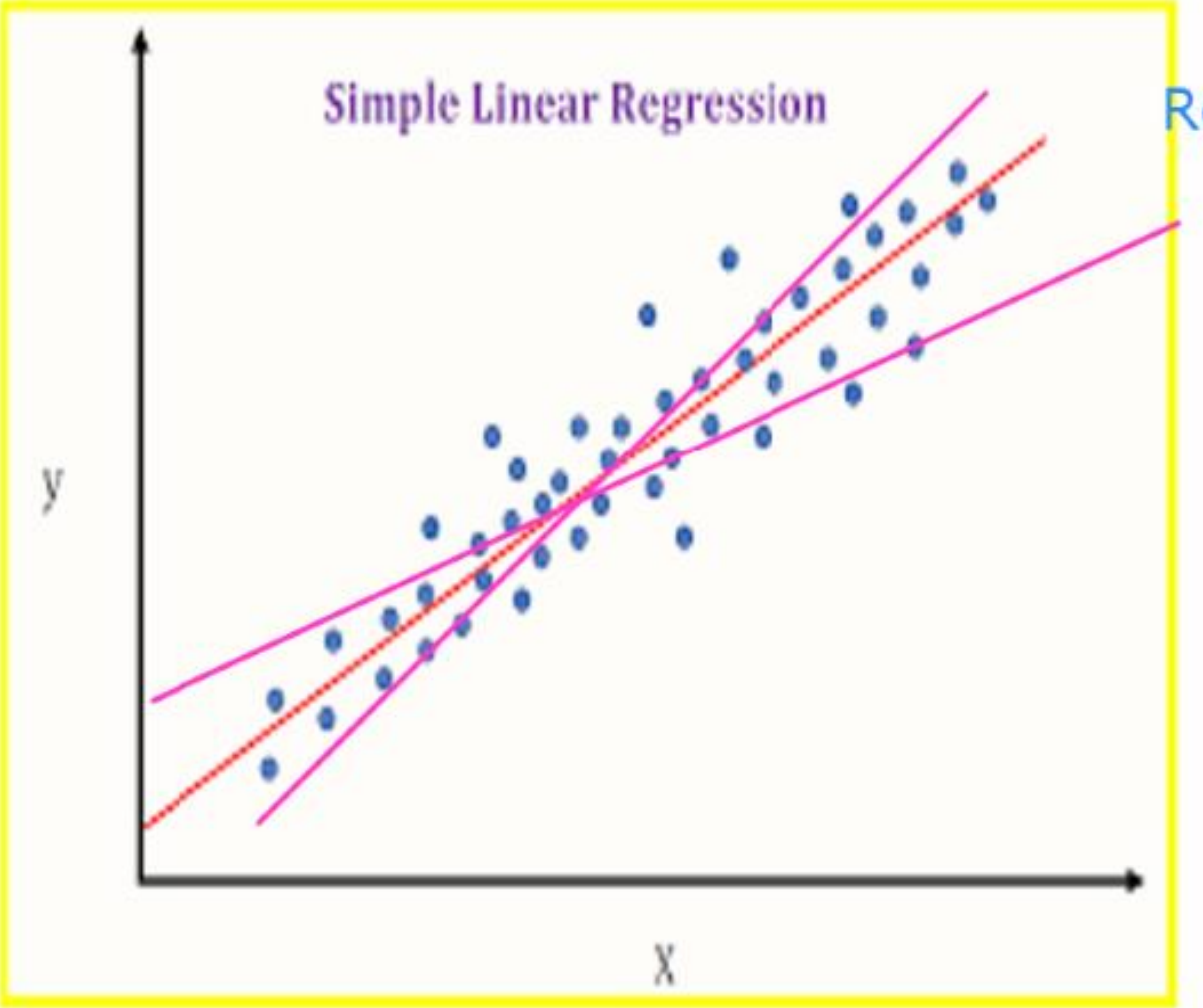
Multiple Linear Regression

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n$$

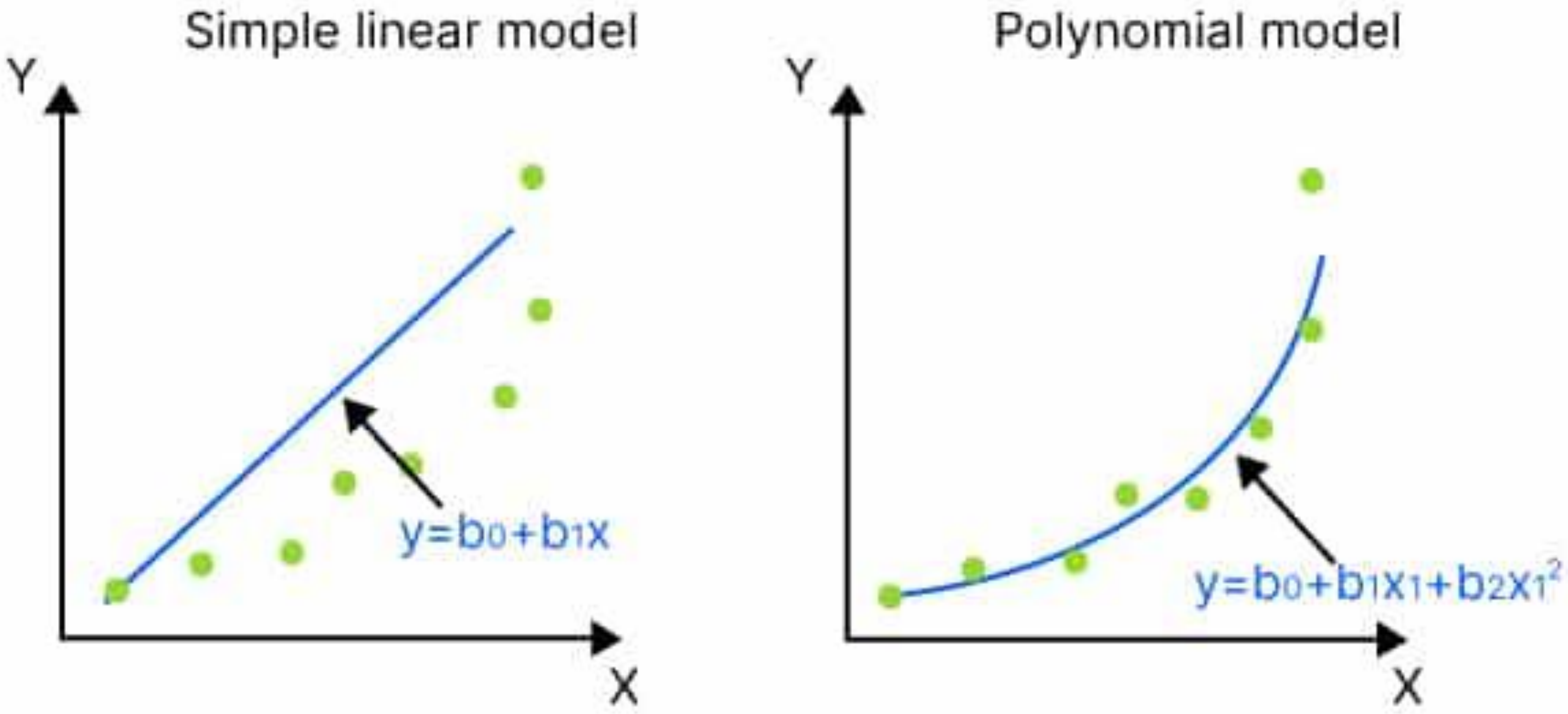
Polynomial Linear Regression

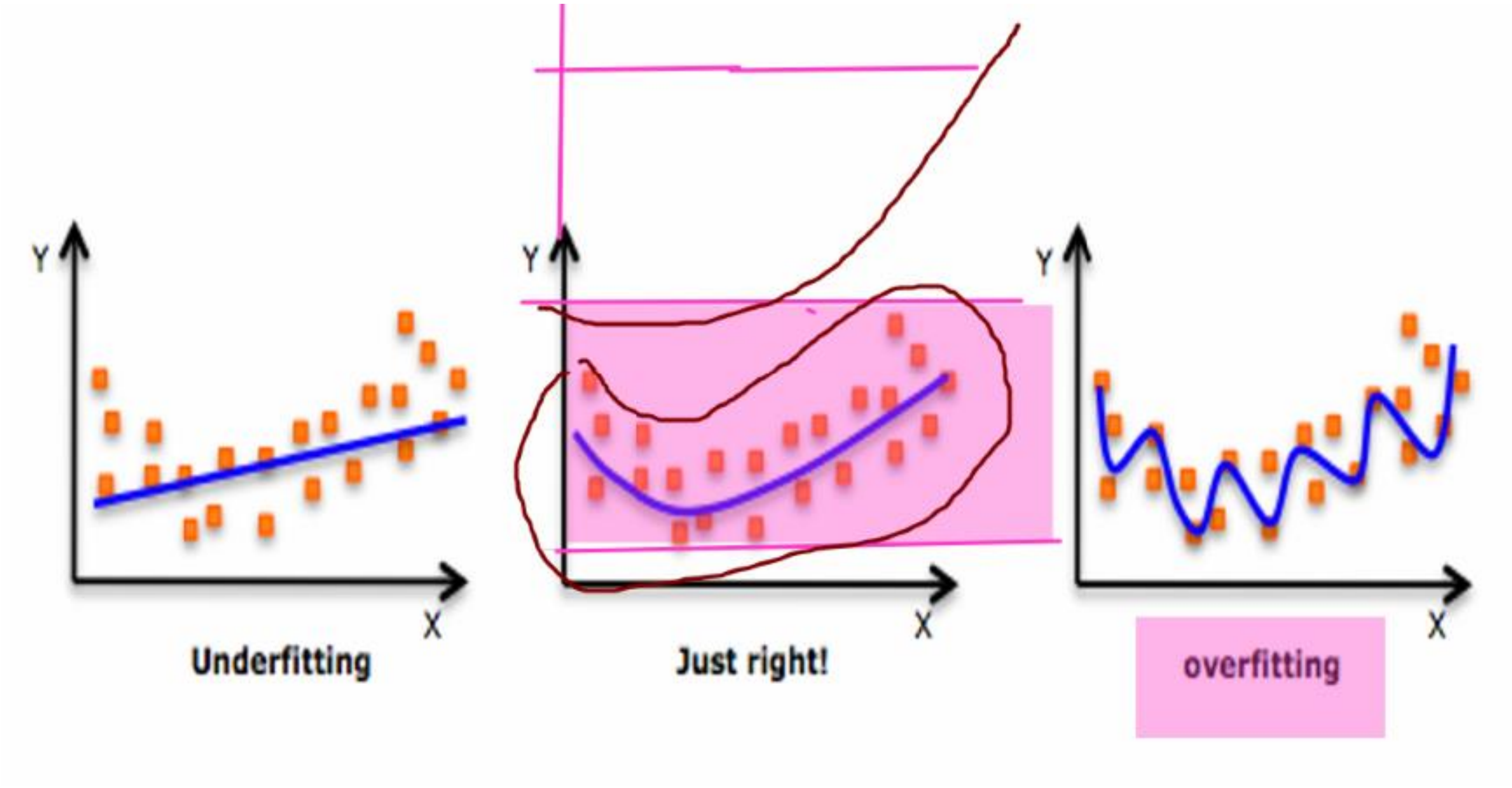
$$y = b_0 + b_1 x_1 + b_2 x_1^2 + \ldots + b_n x_1^n$$

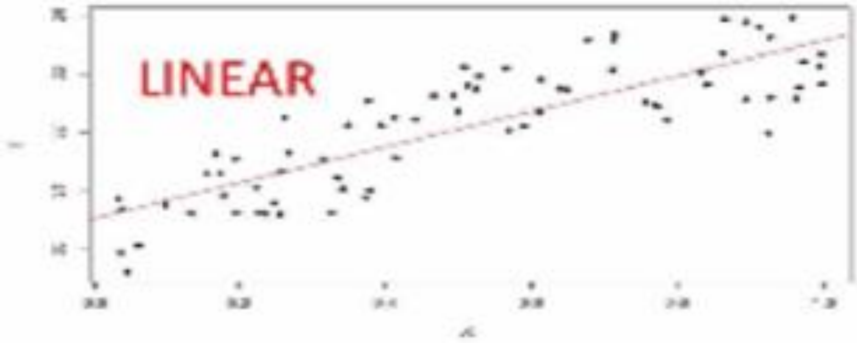
CDAC Mumbai: Kiran Waghmare 3



Regression line(best fit)


Simple linear model
Y
X
y=b0+b1x
Polynomial model
Y
X
y=b0+b1x1+b2x1²


Y
X
Underfitting
Y
X
Just right!
Y
X
overfitting

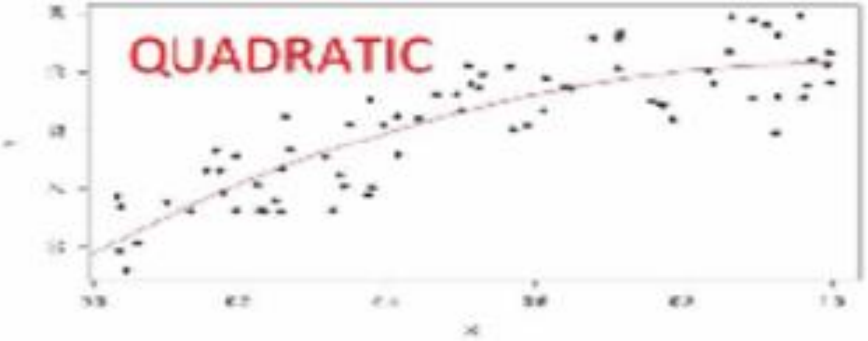


Multiple R-squared: 0.7044

$Y = \boxed{30.53} + \boxed{3.05}*X$

$\boxed{y=2+4x1+8x2}$

2 4 8

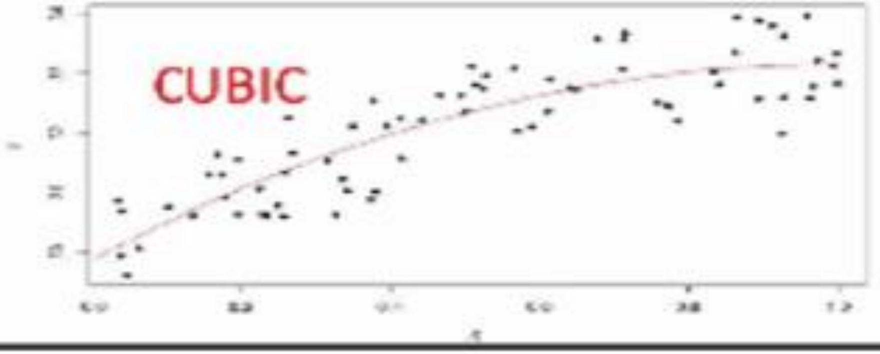


Multiple R-squared: 0.7559

$Y = 29.90 + \boxed{6.48*X} - 3.22*X^2$

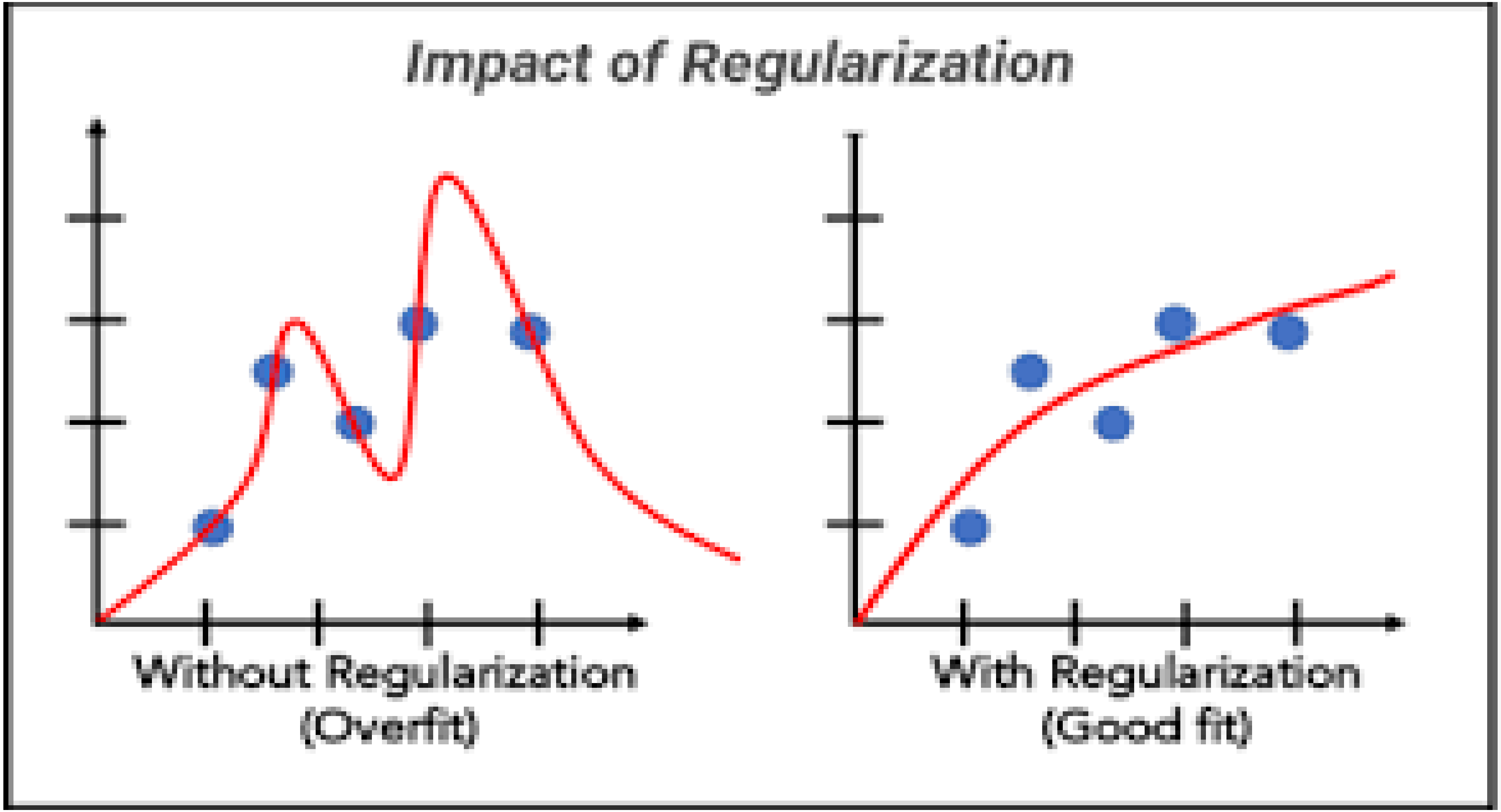
1 2 4

y=1+2x1+4x2



Multiple R-squared: 0.7623

$Y = 30.17 + 3.61*X + 3.71*X^2 - 4.48*X^3$


Impact of Regularization
Without Regularization
(Overfit)
With Regularization
(Good fit)

Transforming the Loss function into Lasso Regression

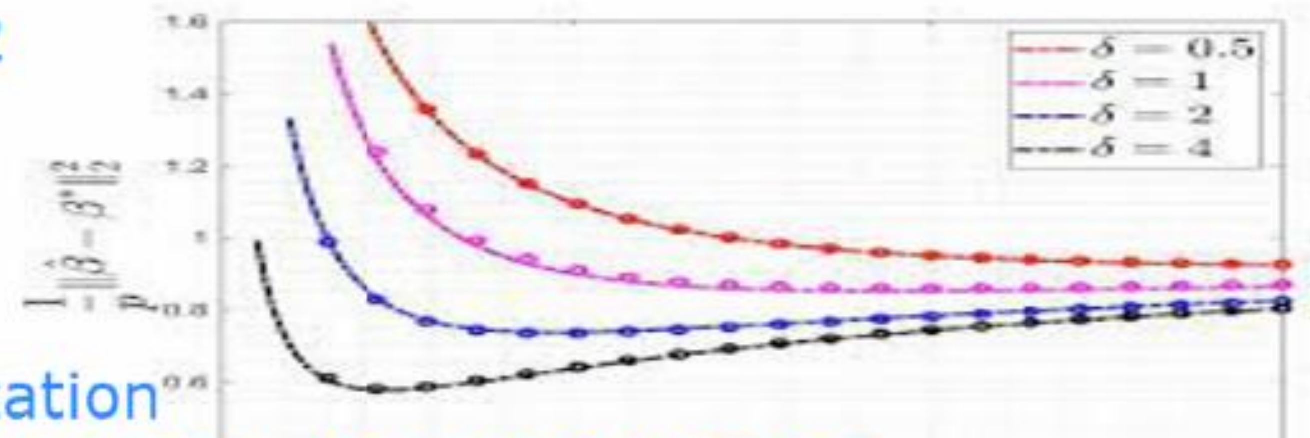
$$\boxed{\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)}^2 \Rightarrow \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{p} w_j \times x_{ij}\right)^2 + \lambda \underbrace{\boxed{\sum_{j=0}^{p}|w_j|}}_{}$$

penalty

Loss function | **Loss function + Regularized term**

Designed by Author (Shanthababu)

y=a+bx1+cx2



Ridge:Regularization
Lasso : Regularization + Feature Selection
Elasticnet: Ridge +Lasso

Ridge Regression

Ridge regression uses the mean squared error loss function and applies L2 Regularization. Its cost function *J(θ)* is given as

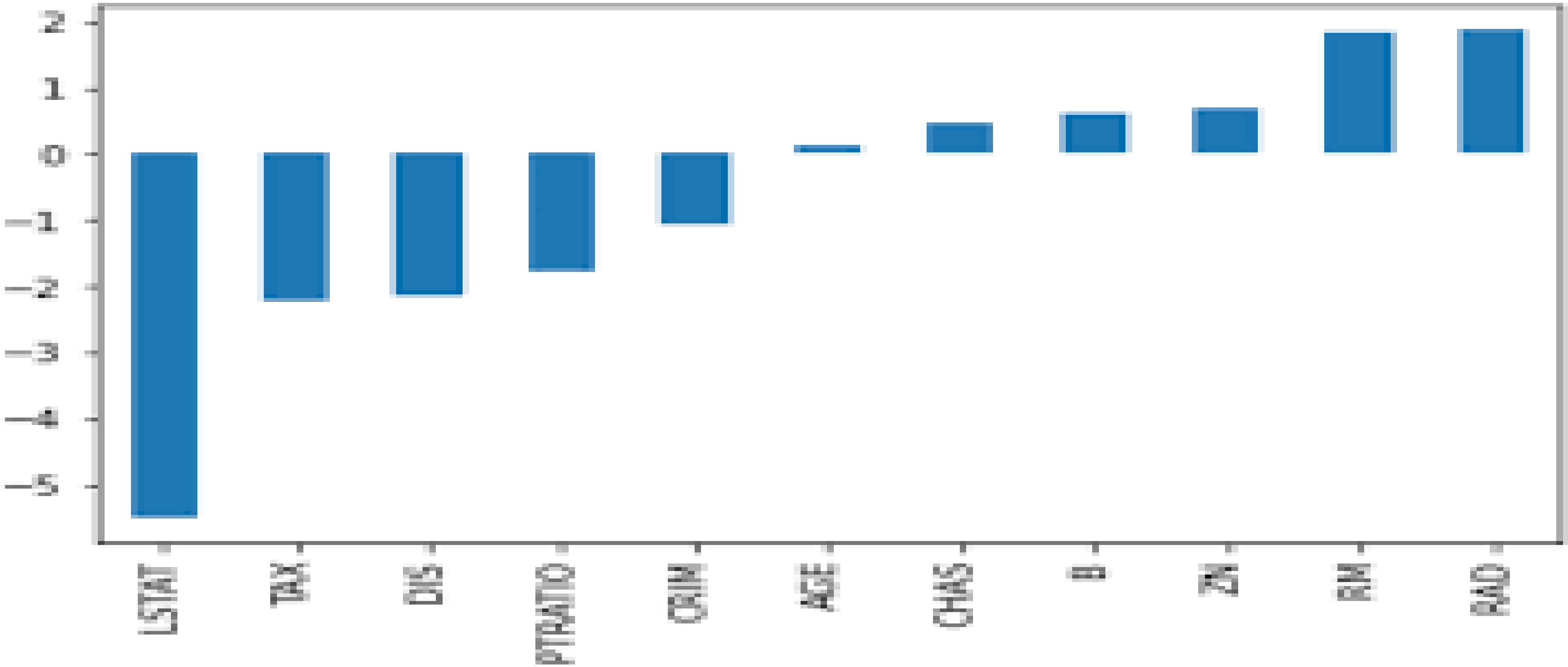
$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}(y - \hat{y})^2 + \lambda \sum_{j=1}^{n} w_j^2$$

where,

$\frac{1}{m}\sum_{i=1}^{m}(y - \hat{y})^2$ is the Mean Squared error (loss function)

$\lambda \sum_{j=1}^{n} w_j^2$ is the penalty (L2 Regularization)

Now, substitute $\hat{y}$ as $wx_i + b$.


2
1
0
−1
−2
−3
−4
−5
LSTAT
TAX
DIS
PTRATIO
CRIM
AGE
CHAS
B
ZN
RM
RAD

Lasso Regression

Lasso regression uses the same mean squared error loss function and this applies L1 Regularization and will repeat the same steps as Ridge. The cost function of Lasso Regression *J(θ)* is given as

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}(y - \hat{y})^2 + \lambda\sum_{j=1}^{n}|w_j|$$

where

$\lambda\sum_{j=1}^{n}|w_j|$ is the penalty (L1 Regularization).

```
In [14]: new_data['Houseprice']=data.target
```

```
In [15]: new_data.head()
```

Out[15]:

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	Houseprice
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1.0	296.0	15.3	396.90	4.98	24.0
1	0.02731	0.0	7.07	0.0	0.469	6.421	78.9	4.9671	2.0	242.0	17.8	396.90	9.14	21.6
2	0.02729	0.0	7.07	0.0	0.469	7.185	61.1	4.9671	2.0	242.0	17.8	392.83	4.03	34.7
3	0.03237	0.0	2.18	0.0	0.458	6.998	45.8	6.0622	3.0	222.0	18.7	394.63	2.94	33.4
4	0.06905	0.0	2.18	0.0	0.458	7.147	54.2	6.0622	3.0	222.0	18.7	396.90	5.33	36.2

```
In [ ]:
```

In [45]:
```
la_coeff
```

Out[45]:
```
CRIM       -0.000000
ZN          0.000000
INDUS      -0.000000
CHAS        0.000000
NOX        -0.000000
RM          2.540098
AGE        -0.000000
DIS        -0.000000
RAD        -0.000000
TAX        -0.171527
PTRATIO    -1.784796
B           0.110959
LSTAT      -3.585324
dtype: float64
```

Alpha

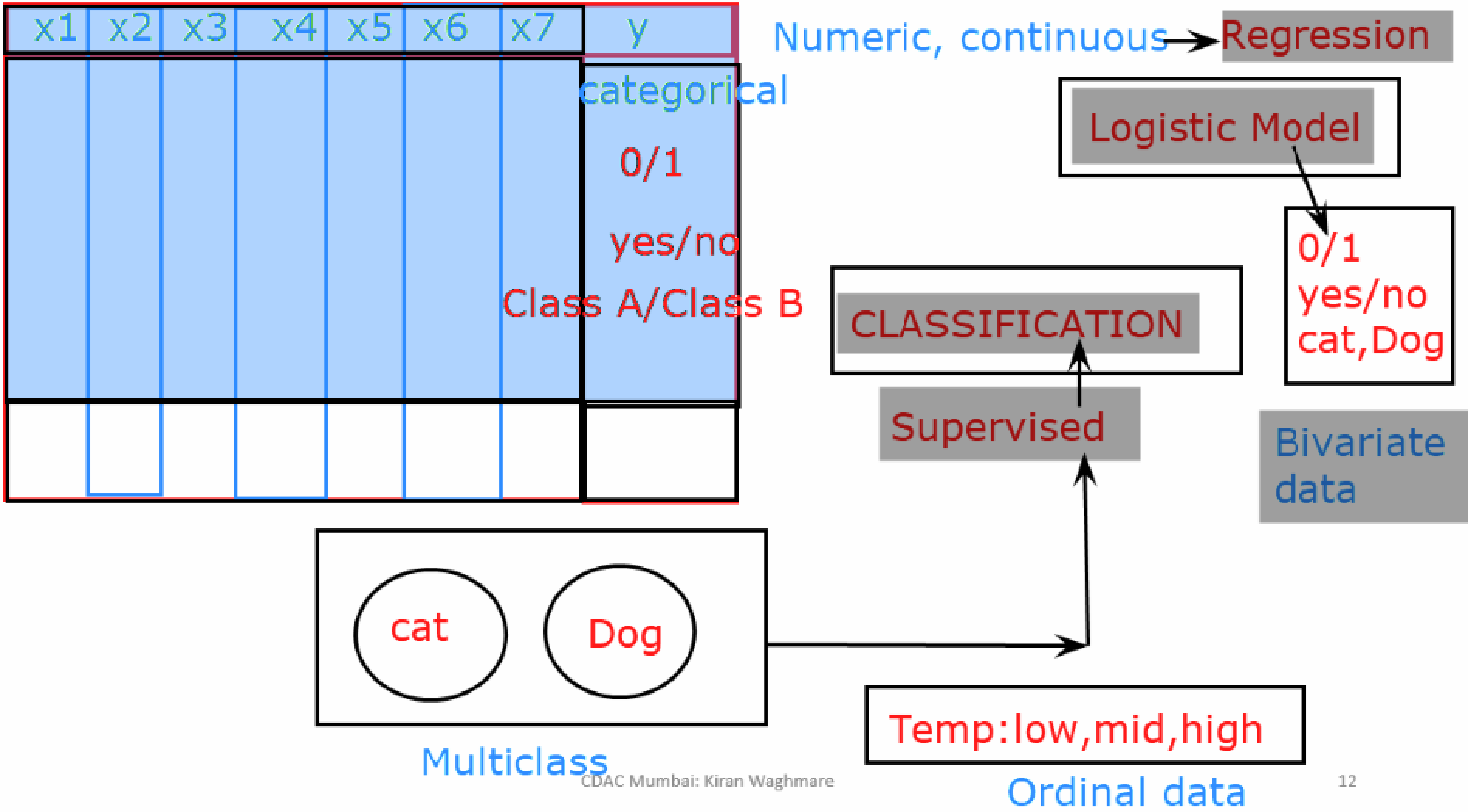
α

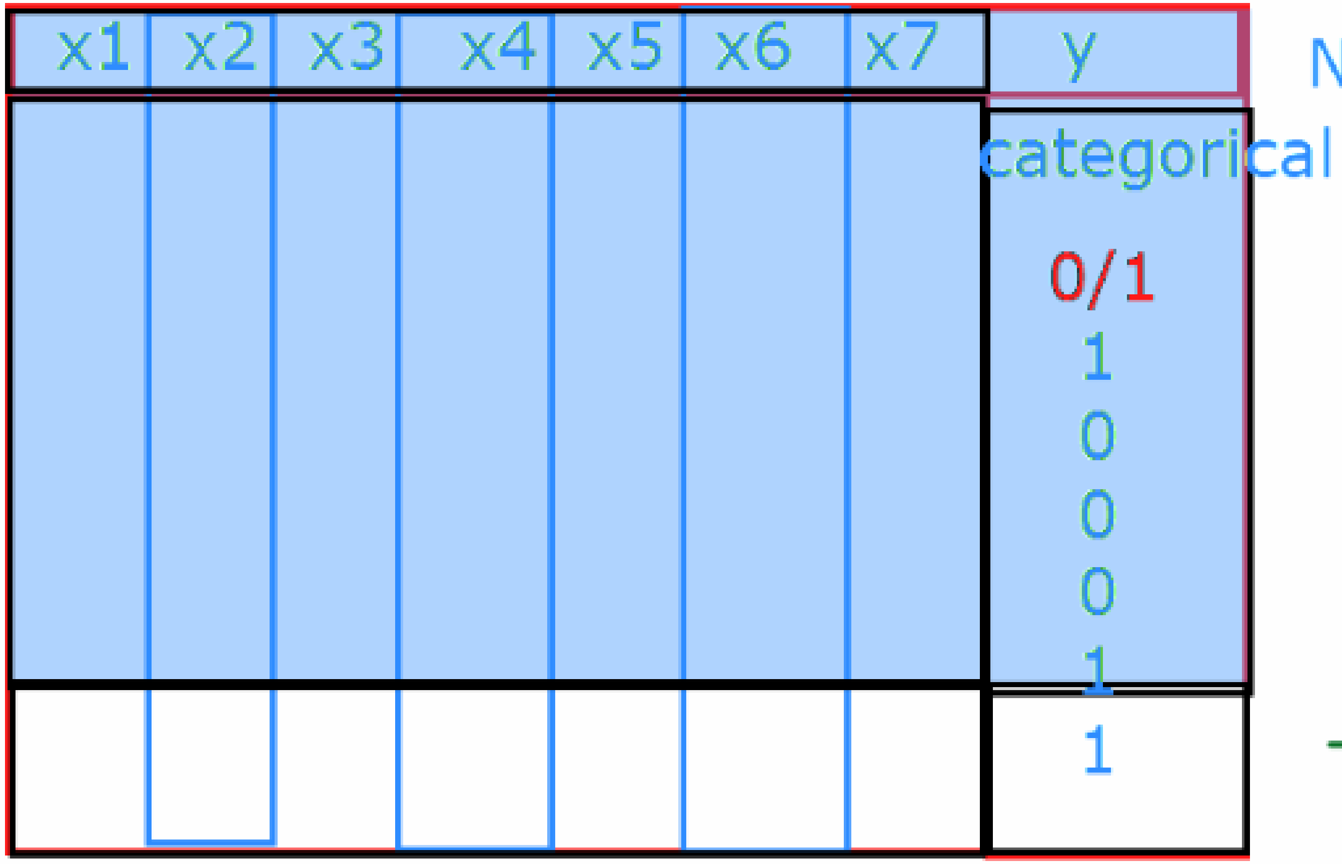
Feature Selection

Regularization

In [46]:
```
la_coeff.plot(kind="barh")
```

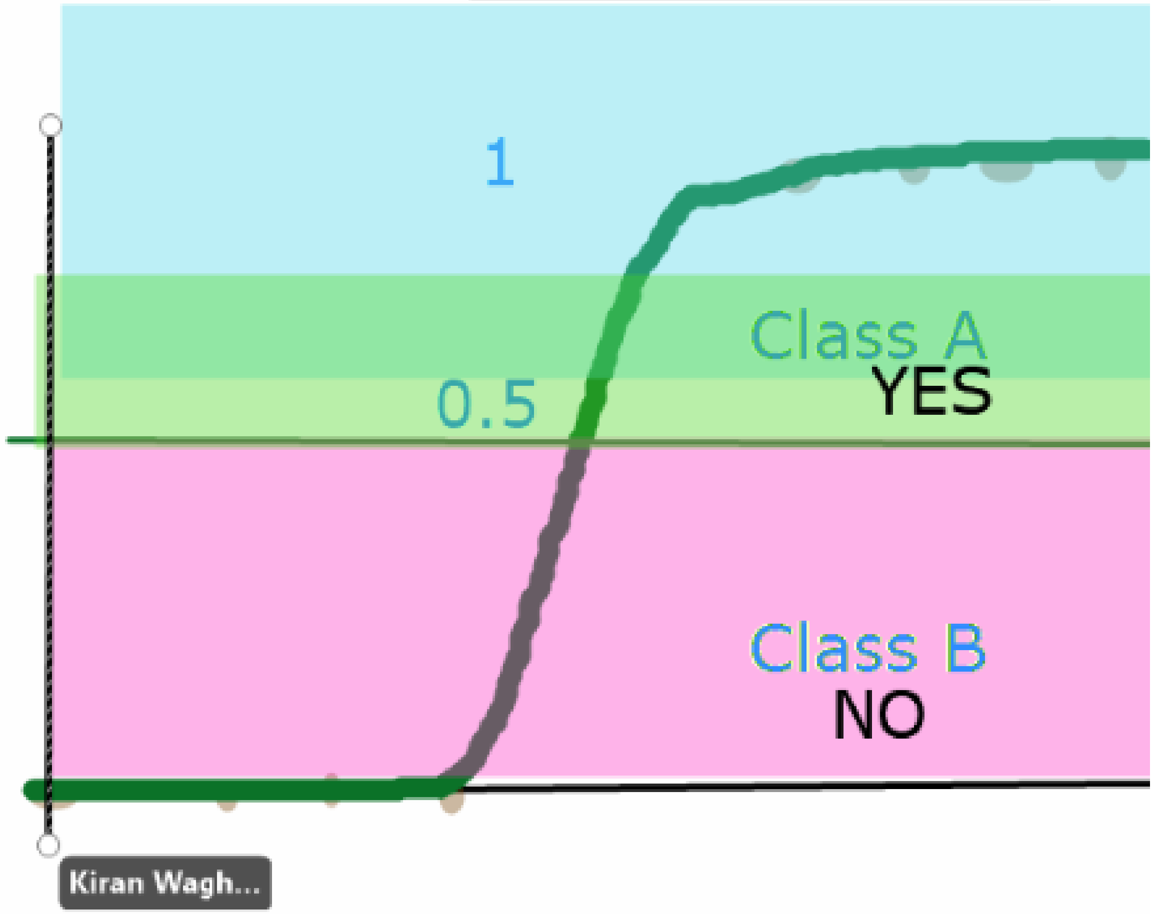
Out[46]:
```
<AxesSubplot:>
```


x1
x2
x3
x4
x5
x6
x7
y
Numeric, continuous
Regression
categorical
0/1
yes/no
Class A/Class B
Logistic Model
0/1
yes/no
cat,Dog
CLASSIFICATION
Supervised
Bivariate data
cat
Dog
Multiclass
Temp:low,mid,high
Ordinal data



Numeric, continuous → Regression

Logistic Model



Supervised Classification

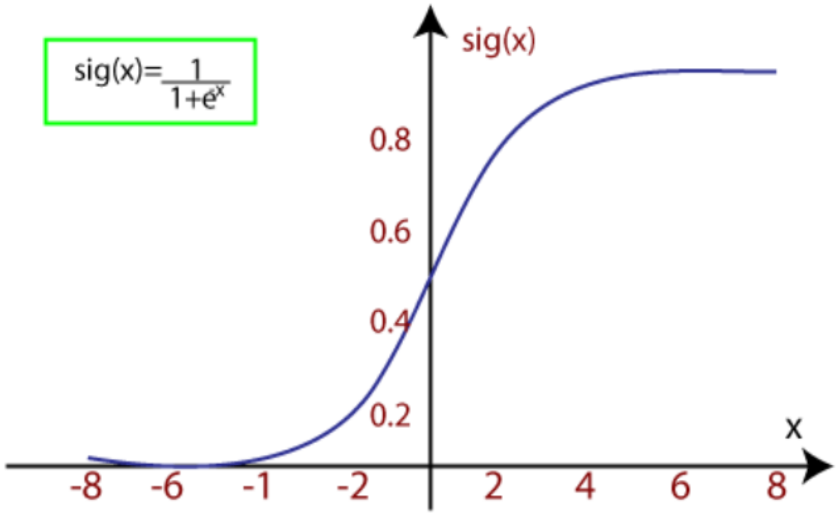
sigmo 0

Logistic Regression

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

- f(x)= Output between the 0 and 1 value.
- x= input to the function
- e= base of natural logarithm.

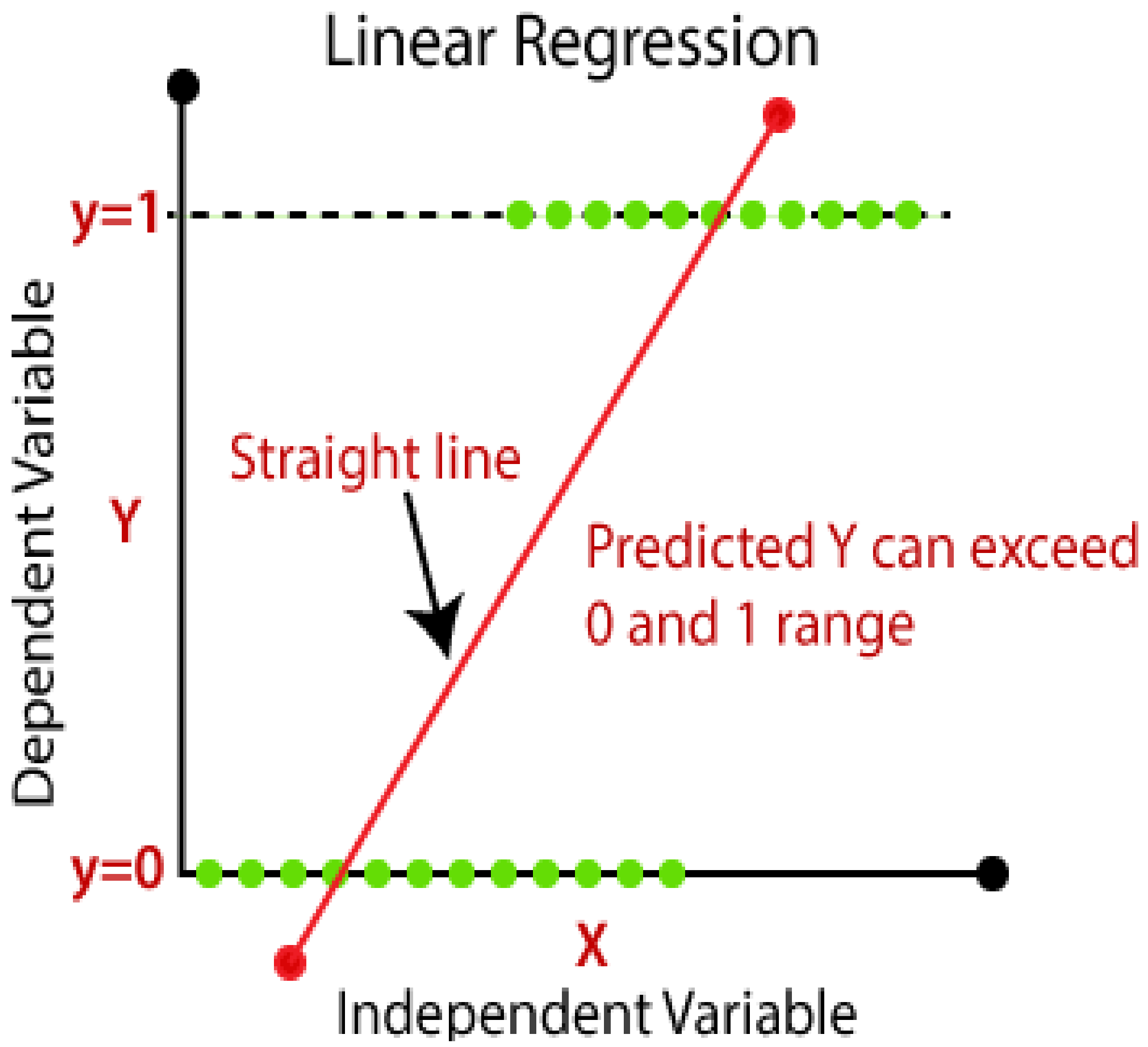
When we provide the input values (data) to the function, it gives the S-curve as follows:

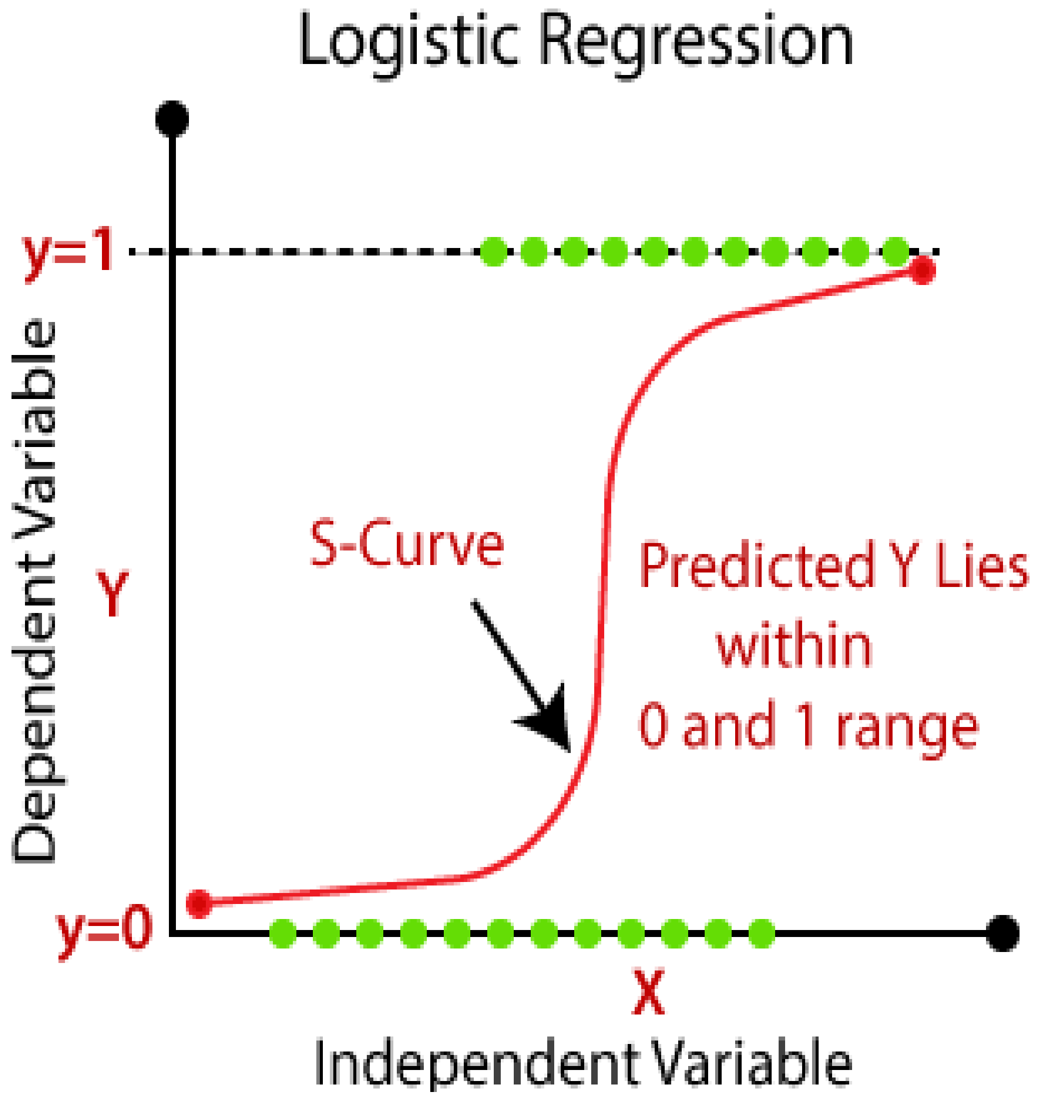


- It uses the concept of threshold levels, values above the threshold level are rounded up to 1, and values below the threshold level are rounded up to 0.

There are three types of logistic regression:

- **Binary(0/1, pass/fail)**
- **Multi(cats, dogs, lions)**
- **Ordinal(low, medium, high)**


Linear Regression
y=1
y=0
Dependent Variable
Y
Straight line
Predicted Y can exceed
0 and 1 range
X
Independent Variable


Logistic Regression
y=1
y=0
Dependent Variable
Y
S-Curve
Predicted Y Lies
within
0 and 1 range
X
Independent Variable