Dataset: 1000 data points

Mini-batch gradient descent algorithm

Dataset: 1000 data points

Mini-batch gradient descent algorithm

2, 4, 8, 16, 32, 64, 128, so on

Dataset: 1000 data points

Mini-batch gradient descent algorithm

2, 4, 8, 16, 32, 64, 128, so on

128: batch size

Dataset: 1000 data points

Mini-batch gradient descent algorithm

2, 4, 8, 16, 32, 64, 128, so on

128: batch size

896: 2^7 : 104 batch 8

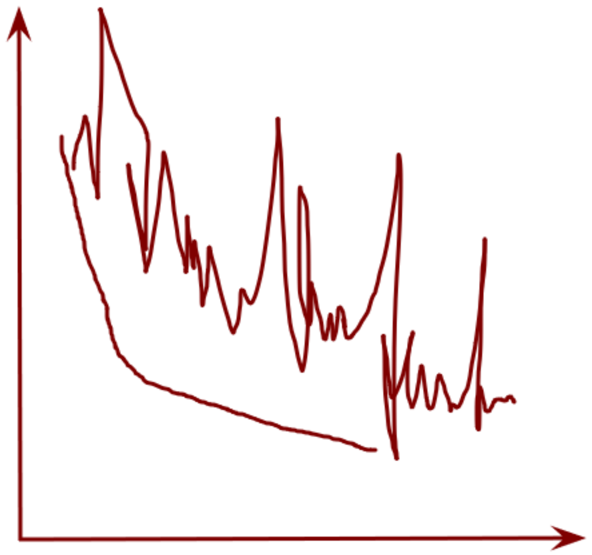
Dataset: 1000 data points

Mini-batch gradient descent algorithm

2, 4, 8, 16, 32, 64, 128, so on

128: batch size

896: 2^7 : 104 batch 8



Dataset: 1000 data points

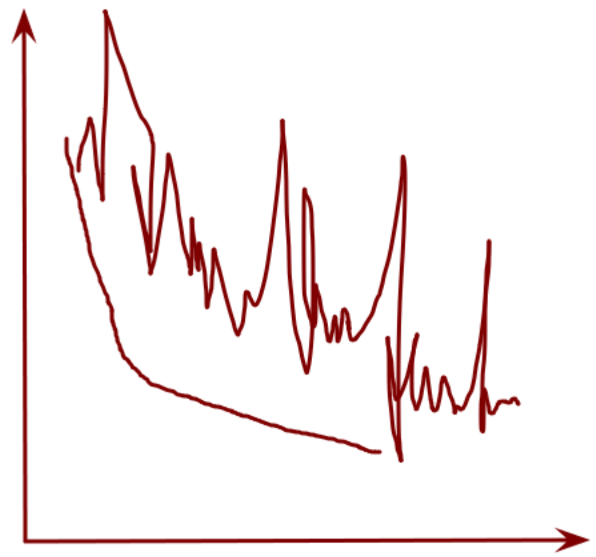
Mini-batch gradient descent algorithm

2, 4, 8, 16, 32, 64, 128, so on

128: batch size

896: 2^7 : 104 batch 8

1000-3000 data points



Advantages & Disadvantages of Batch Gradient Descent Algo

Advantages

- It is computationally efficient
- It has stable performance (less noise)

Disadvantages

- It requires a lot of memory
- It has a slower learning process
- It may become caught in local minima

Advantages & Disadvantages of Stochastic Gradient Descent Algo

Advantages

- It has faster learning on some problems
- The algorithm is simple to understand
- It provides immediate feedback

Disadvantages

- It is computationally intensive
- There’s a definite probability it won’t settle in the global minimum
- The performance will be very noisy

Advantages & Disadvantages of Mini Batch Gradient Descent Algo

Advantages

- It avoids getting stuck in local minima
- It is more computationally efficient than Stochastic Gradient Descent (SGD)
- It does not need as much memory as Batch Gradient Descent (BGD)

Disadvantages

- Hyperparameter Tuning (batch_size)
- It can be highly expensive and intractable for datasets that are too large to fit in memory.

Logistic Regression

Classification:
Email: Spam/no spam
online transactions: fradulent/not
Tumor: Malignant/Begin
Diabetic: Yes/No
Y belongs to either 0 or 1
0: negative class
1: positive class
Binary classification problems: two ouputs
Multi-class classification problems: multiple outputs {0, 1, 2, 3, 4, 5, 6, ...}

Classification:
Email: Spam/no spam
online transactions: fradulent/not
Tumor: Malignant/Begin
Diabetic: Yes/No
Y belongs to either 0 or 1
0: negative class
1: positive class
Binary classification problems: two ouputs
Multi-class classification problems: multiple outputs {0, 1, 2, 3, 4, 5, 6, ...}

Y = f(x): H(x)= theta sub 0 + theta sub 1 x1 + theta sub 2 x2 +........+ theta sub n xn)

Classification:
Email: Spam/no spam
online transactions: fradulent/not
Tumor: Malignant/Begin
Diabetic: Yes/No
Y belongs to either 0 or 1
0: negative class
1: positive class
Binary classification problems: two ouputs
Multi-class classification problems: multiple outputs {0, 1, 2, 3, 4, 5, 6, ...}

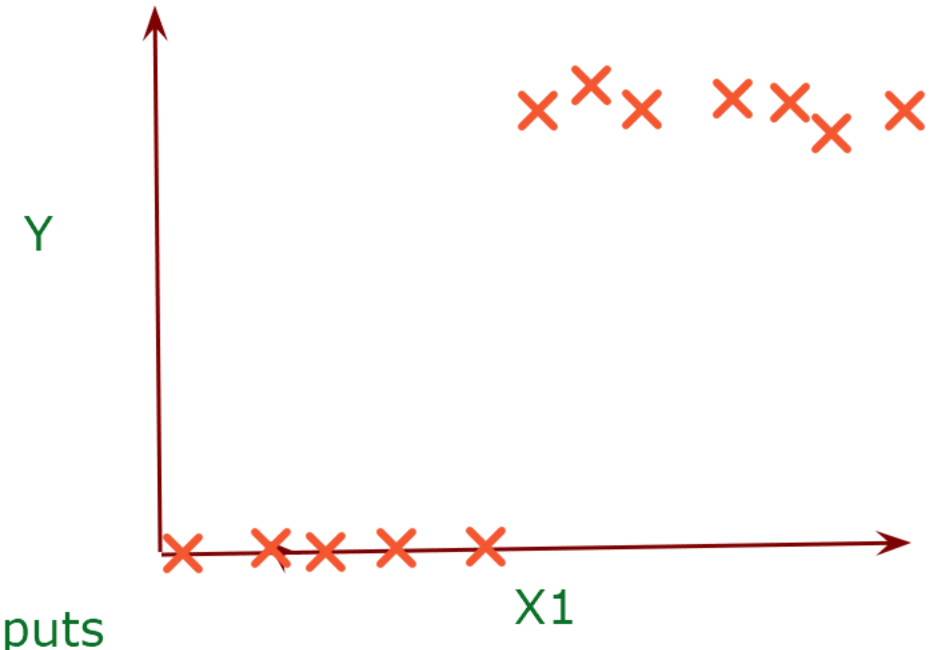
Y = f(x): H(x)= theta sub 0 + theta sub 1 x1 + theta sub 2 x2 +........+ theta sub n xn)

0<= H(x) <= 1

Classification:
Email: Spam/no spam
online transactions: fradulent/not
Tumor: Malignant/Begin
Diabetic: Yes/No
Y belongs to either 0 or 1
0: negative class
1: positive class
Binary classification problems: two ouputs
Multi-class classification problems: multiple outputs {0, 1, 2, 3, 4, 5, 6, ...}

Y = f(x): H(x)= theta sub 0 + theta sub 1 x1 + theta sub 2 x2 +.......+ theta sub n xn)

0<= H(x) <= 1

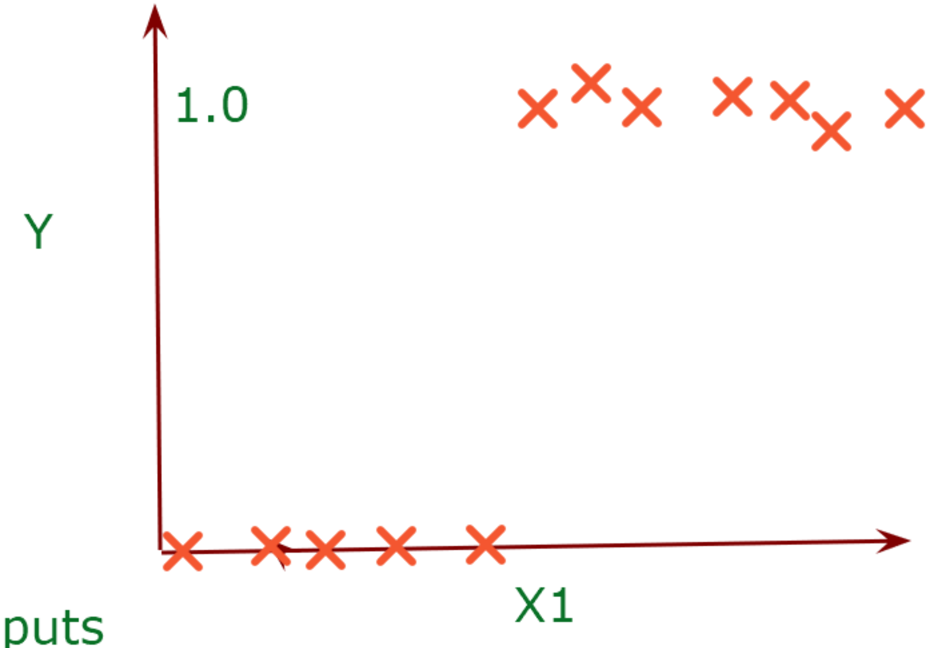


Classification:
Email: Spam/no spam
online transactions: fradulent/not
Tumor: Malignant/Begin
Diabetic: Yes/No
Y belongs to either 0 or 1
0: negative class
1: positive class
Binary classification problems: two ouputs
Multi-class classification problems: multiple outputs {0, 1, 2, 3, 4, 5, 6, ...}

Y = f(x): H(x)= theta sub 0 + theta sub 1 x1 + theta sub 2 x2 +.......+ theta sub n xn)

$0 <= H(x) <= 1$

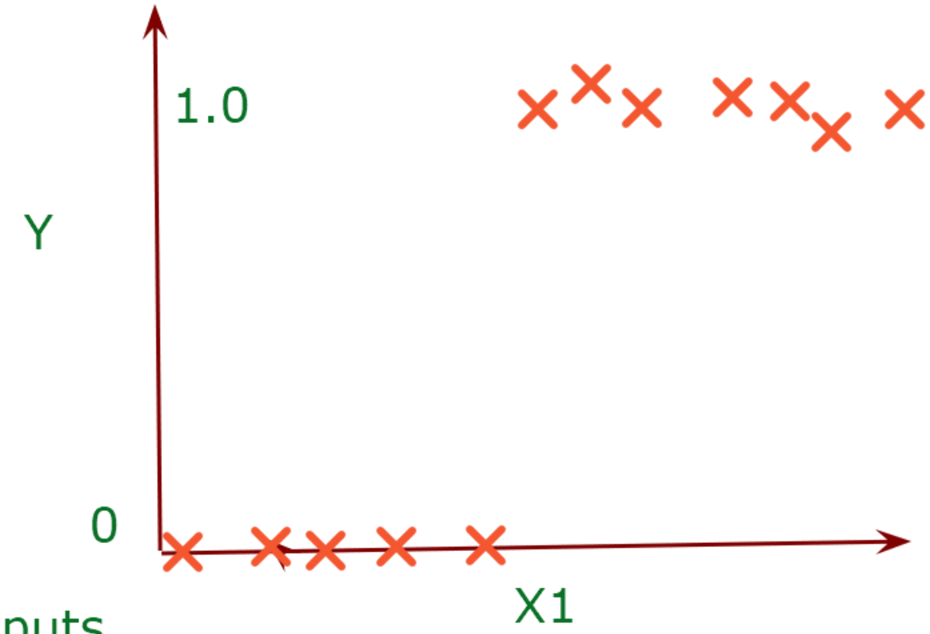
Classification:
Email: Spam/no spam
online transactions: fradulent/not
Tumor: Malignant/Benign
Diabetic: Yes/No
Y belongs to either 0 or 1
0: negative class
1: positive class
Binary classification problems: two ouputs
Multi-class classification problems: multiple outputs {0, 1, 2, 3, 4, 5, 6, ...}



Y = f(x): H(x)= theta sub 0 + theta sub 1 x1 + theta sub 2 x2 +.......+ theta sub n xn)

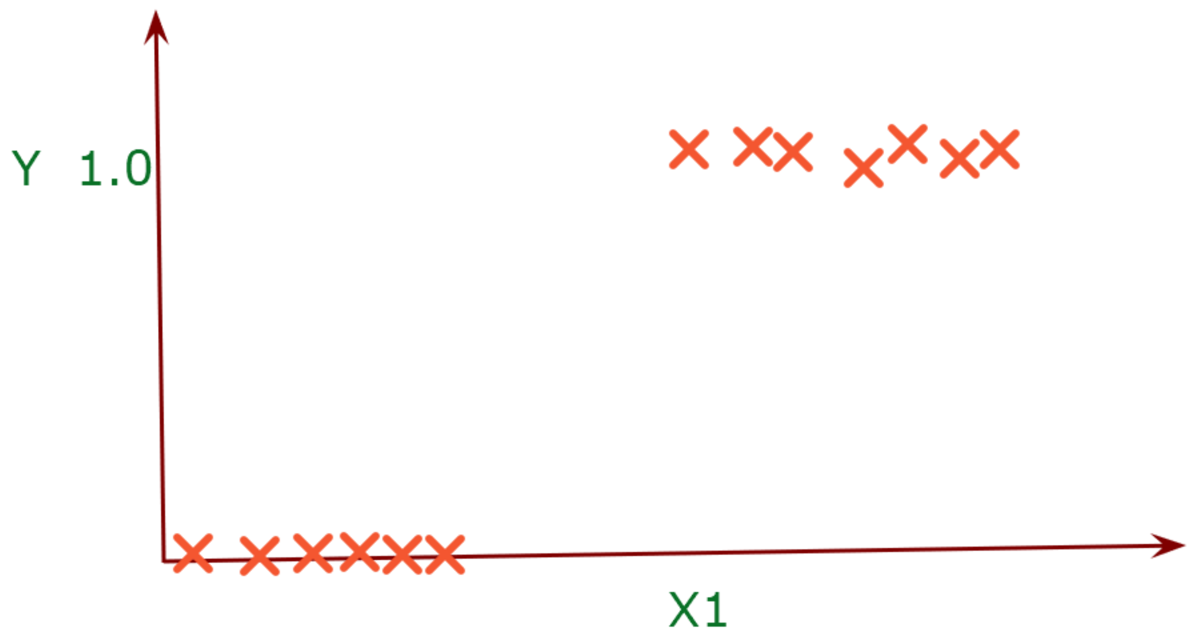
0<= H(x) <= 1

Classification:
Email: Spam/no spam
online transactions: fradulent/not
Tumor: Malignant/Benign
Diabetic: Yes/No
Y belongs to either 0 or 1
0: negative class
1: positive class
Binary classification problems: two ouputs
Multi-class classification problems: multiple outputs {0, 1, 2, 3, 4, 5, 6, ...}



Y = f(x): H(x)= theta sub 0 + theta sub 1 x1 + theta sub 2 x2 +........+ theta sub n xn)

0<= H(x) <= 1


Y 1.0
X1

Logistic Regression

Classification: Y = discrete in nature 0 or 1

Logistic Regression

Classification: Y = discrete in nature 0 or 1

Y belongs to either yes or No

Logistic Regression

Classification: Y = discrete in nature 0 or 1

Y belongs to either yes or No

predictive analysis

Logistic Regression

Classification: Y = discrete in nature 0 or 1

Y belongs to either yes or No

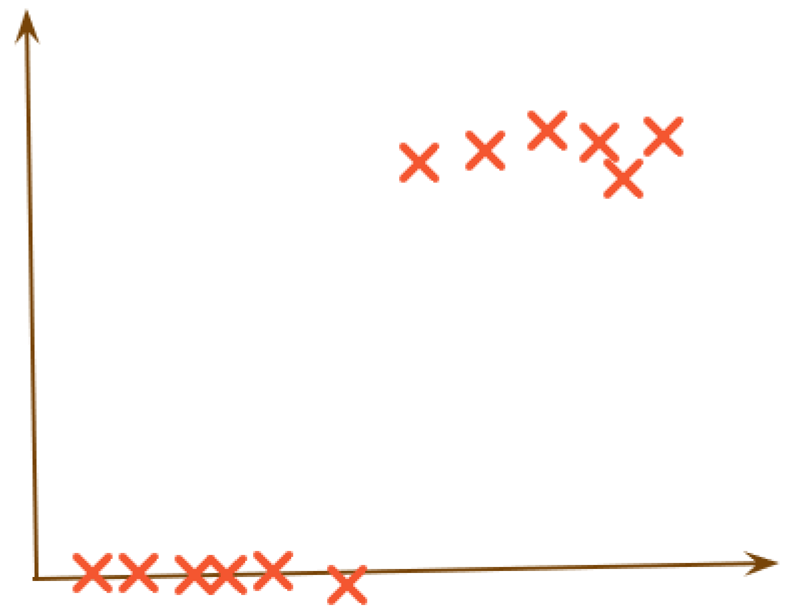
predictive analysis

Logistic Regression

Classification: Y = discrete in nature 0 or 1

Y belongs to either yes or No

predictive analysis

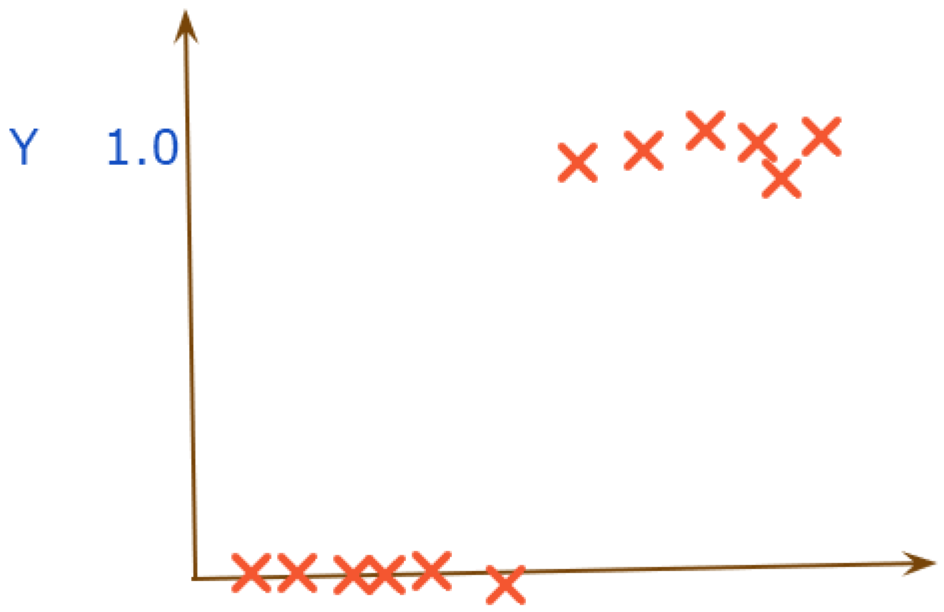


Logistic Regression

Classification: Y = discrete in nature 0 or 1

Y belongs to either yes or No

predictive analysis

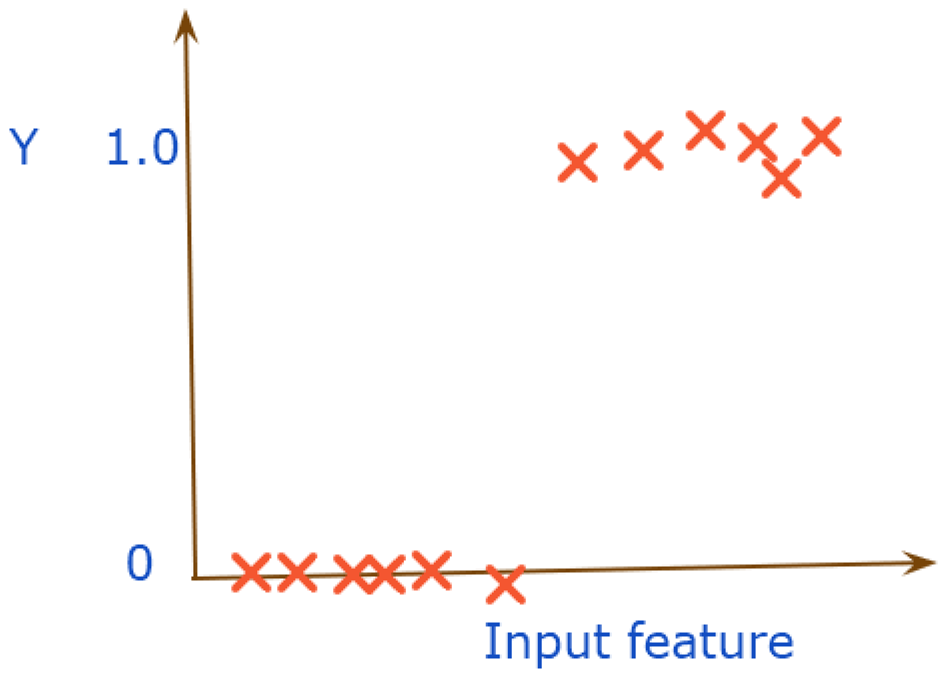


Logistic Regression

Classification: Y = discrete in nature 0 or 1

Y belongs to either yes or No

predictive analysis

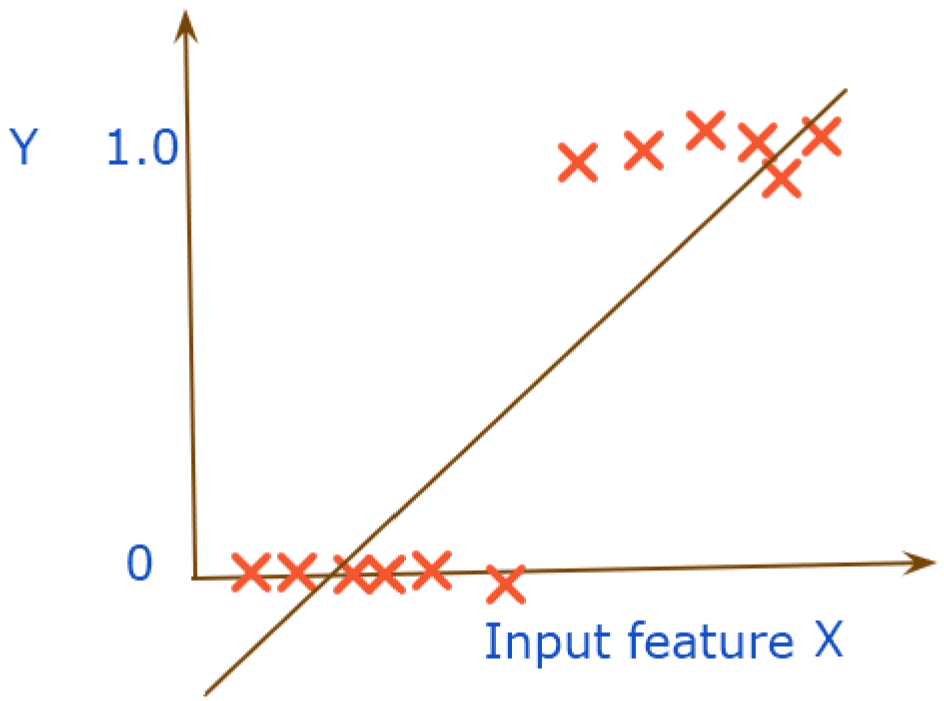


Logistic Regression

Classification: Y = discrete in nature 0 or 1

Y belongs to either yes or No

predictive analysis



Logistic Regression

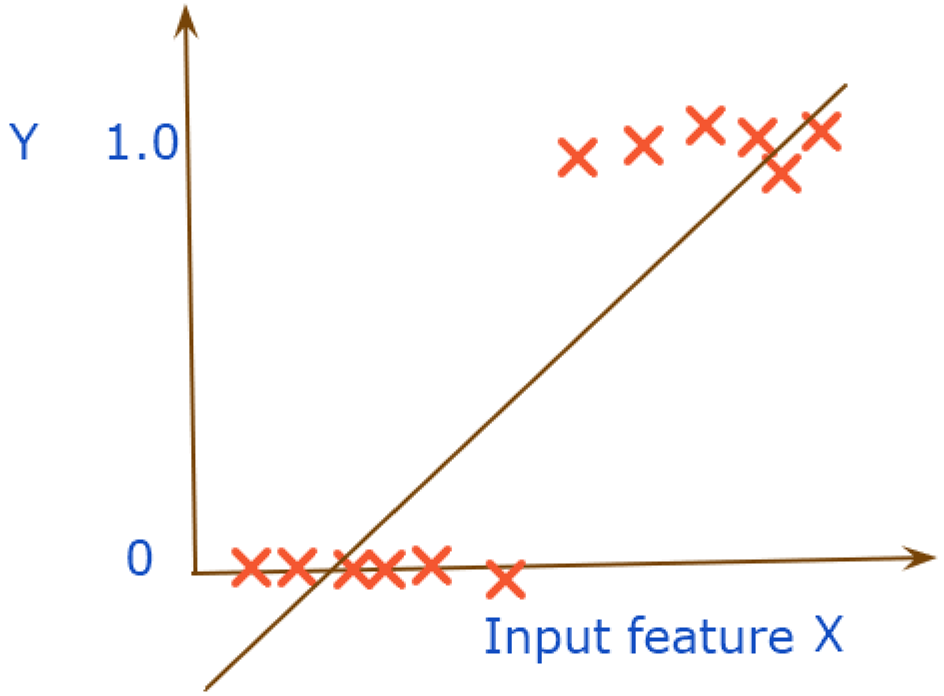
Classification: Y = discrete in nature 0 or 1

Y belongs to either yes or No

predictive analysis

threshold value

H(x) = theta sub 0 + theta sub 1 (X)



Logistic Regression

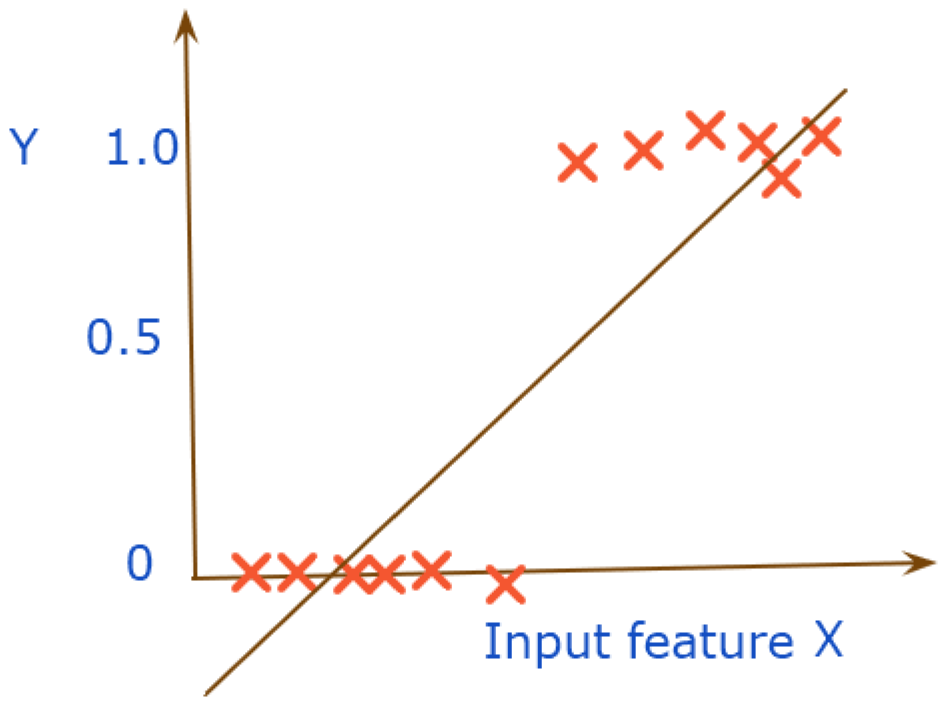
Classification: Y = discrete in nature 0 or 1

Y belongs to either yes or No

predictive analysis

threshold value

H(x) = theta sub 0 + theta sub 1 (X)



Logistic Regression

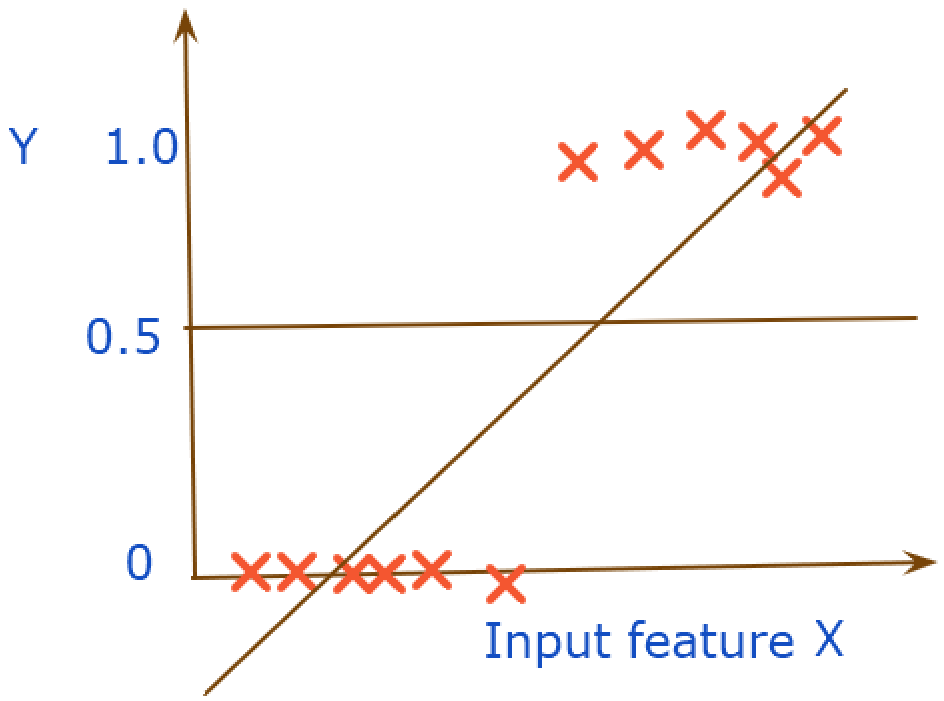
Classification: Y = discrete in nature 0 or 1

Y belongs to either yes or No

predictive analysis

threshold value

H(x) = theta sub 0 + theta sub 1 (X)



Logistic Regression

Classification: Y = discrete in nature 0 or 1

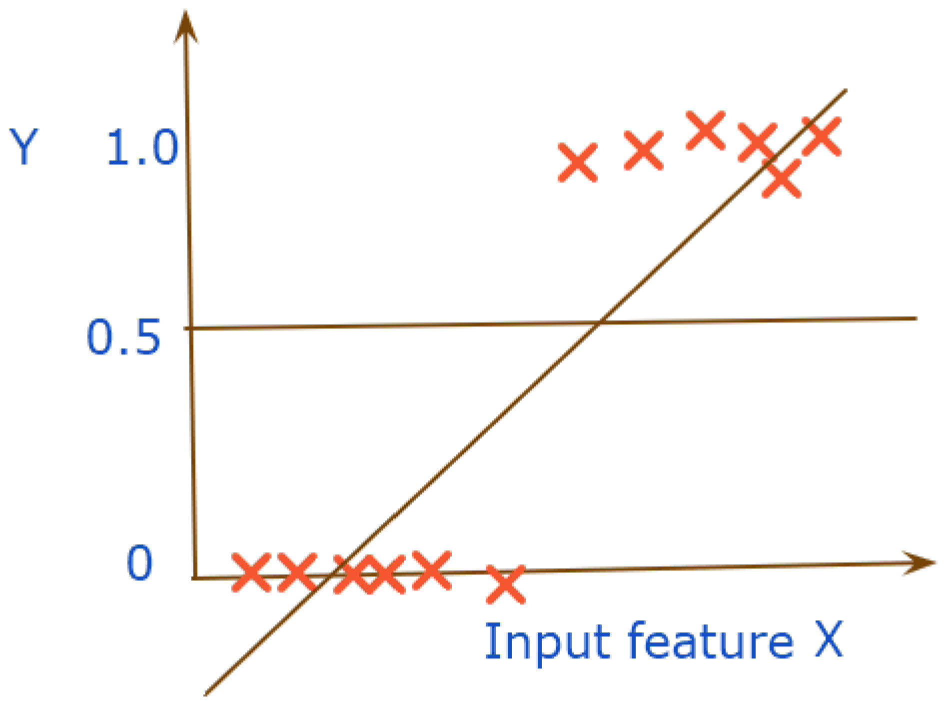
Y belongs to either yes or No

predictive analysis

threshold value

H(x) = theta sub 0 + theta sub 1 (X)

H(x) < 0.5; predict y = 0
H(x) >= 0.5; predict y = 1



Logistic Regression

Classification: Y = discrete in nature 0 or 1

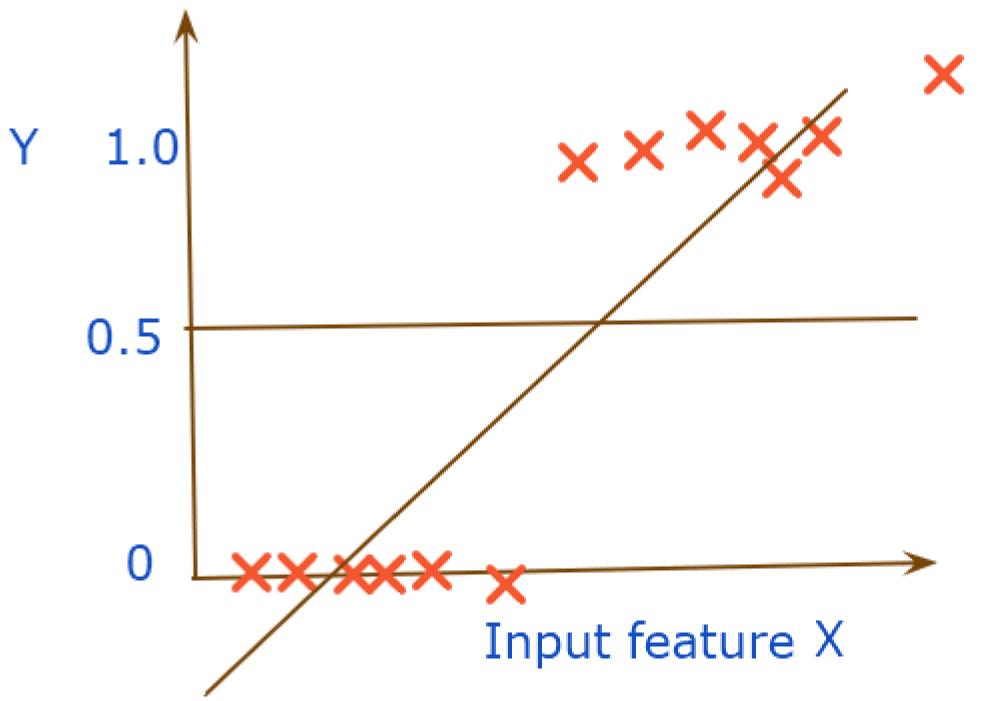
Y belongs to either yes or No

predictive analysis

threshold value

H(x) = theta sub 0 + theta sub 1 (X)

H(x) < 0.5; predict y = 0
H(x) >= 0.5; predict y = 1



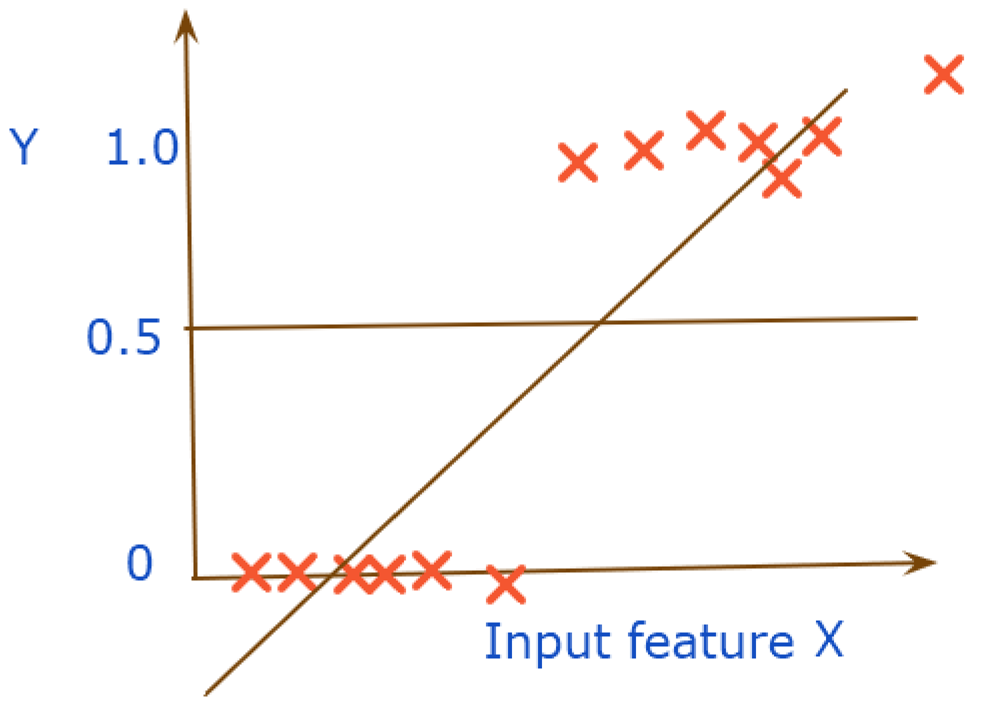
Logistic Regression

Classification: Y = discrete in nature 0 or 1

Y belongs to either yes or No

predictive analysis

threshold value



H(x) = theta sub 0 + theta sub 1 (X)

H(x) < 0.5; predict y = 0
H(x) >= 0.5; predict y = 1

Y = 3.5 lakhs for given size of house
Y = 0 given input feature x1

Logistic Regression

Classification: Y = discrete in nature 0 or 1

Y belongs to either yes or No

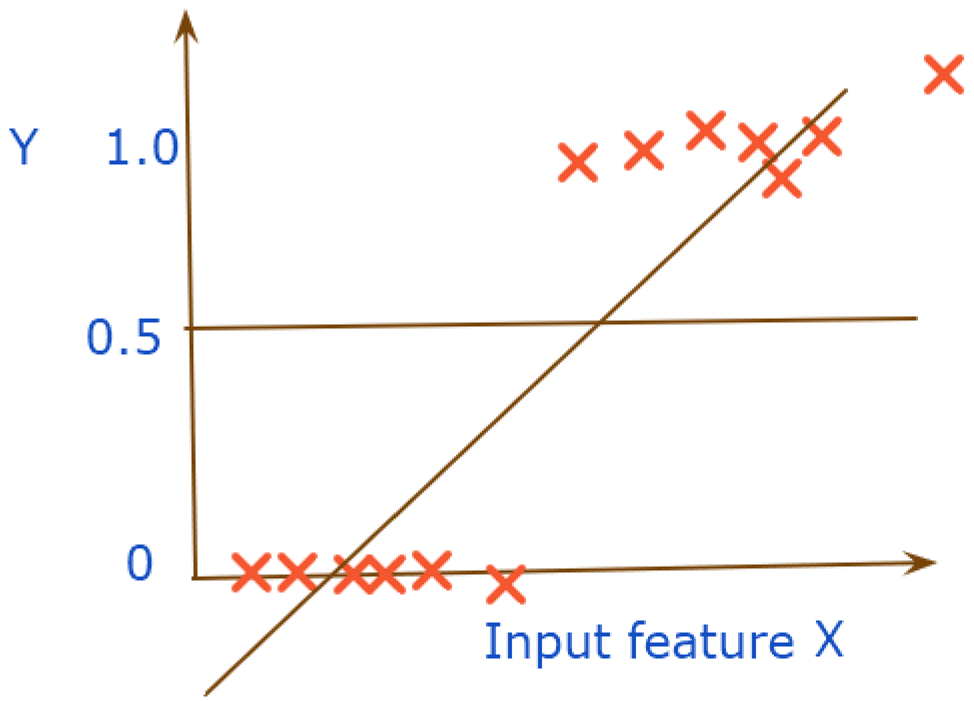
predictive analysis

threshold value

H(x) = theta sub 0 + theta sub 1 (X)

H(x) < 0.5; predict y = 0
H(x) >= 0.5; predict y = 1

Y = 3.5 lakhs for given size of house
Y = 0 given input feature x1



predict the probabilities

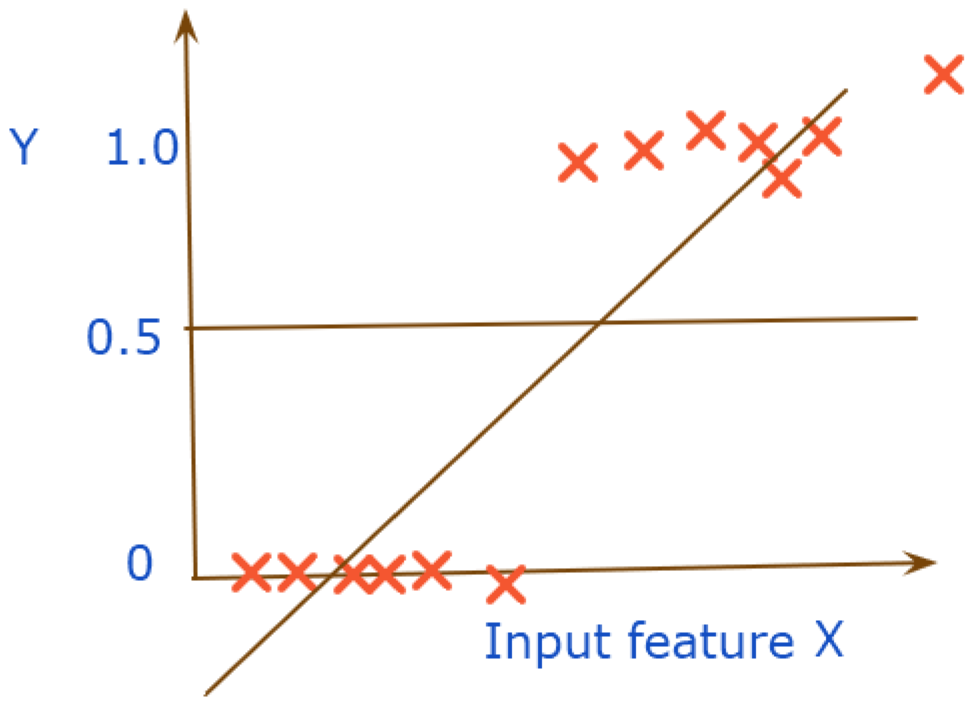
Logistic Regression

Classification: Y = discrete in nature 0 or 1

Y belongs to either yes or No

predictive analysis

threshold value



H(x) = theta sub 0 + theta sub 1 (X)

H(x) < 0.5; predict y = 0
H(x) >= 0.5; predict y = 1

Y = 3.5 lakhs for given size of house
Y = 0 given input feature x1

predict the probabilities

H(x) = P(y = 1|x; theta)
= P(y = 0|x; theta)

H(x) = y = theta sub 0 + theta sub 1 (x)

H(x) = y = theta sub 0 + theta sub 1 (x)

P = theta sub 0 + theta sub 1 (x)

H(x) = y = theta sub 0 + theta sub 1 (x)

P = theta sub 0 + theta sub 1 (x)

odds of P
(P/(1-P)) = theta sub 0 + theta sub 1 (x)

The odds of an event represent the ratio of the (probability that the event will occur) / (probability that the event will not occur)

H(x) = y = theta sub 0 + theta sub 1 (x)

P = theta sub 0 + theta sub 1 (x)

odds of P
(P/(1-P)) = theta sub 0 + theta sub 1 (x)   (0 to +infinity)

(P/1-P) = theta sub 0 + theta sub 1 (x)
restriction in the range - restrict the no of data points

(P/1-P) = theta sub 0 + theta sub 1 (x)
restriction in the range - restrict the no of data points

log of odds of P = log (P/1-P)

(P/1-P) = theta sub 0 + theta sub 1 (x)
restriction in the range - restrict the no of data points

log of odds of P = log (P/1-P) = theta sub 0 + theta sub 1 (x)

(P/1-P) = theta sub 0 + theta sub 1 (x)
restriction in the range - restrict the no of data points

log of odds of P = log (P/1-P) = theta sub 0 + theta sub 1 (x)

exp[log (P/1-P) = exp [theta sub 0 + theta sub 1 (x)]

(P/1-P) = theta sub 0 + theta sub 1 (x)
restriction in the range - restrict the no of data points

log of odds of P = log (P/1-P) = theta sub 0 + theta sub 1 (x)

exp[log (P/1-P) = exp [theta sub 0 + theta sub 1 (x)]

(p/1-p) = exp [theta sub 0 + theta sub 1 (x)]

(P/1-P) = theta sub 0 + theta sub 1 (x)
restriction in the range - restrict the no of data points

log of odds of P = log (P/1-P) = theta sub 0 + theta sub 1 (x)

exp[log (P/1-P) = exp [theta sub 0 + theta sub 1 (x)]

(p/1-p) = exp [theta sub 0 + theta sub 1 (x)]

P = exp(theta sub 0 + theta sub 1 (x)) - pexp(theta sub 0 + theta sub 1(x)

(P/1-P) = theta sub 0 + theta sub 1 (x)
restriction in the range - restrict the no of data points

log of odds of P = log (P/1-P) = theta sub 0 + theta sub 1 (x)

exp[log (P/1-P) = exp [theta sub 0 + theta sub 1 (x)]

(p/1-p) = exp [theta sub 0 + theta sub 1 (x)]

P = exp(theta sub 0 + theta sub 1 (x)) - pexp(theta sub 0 + theta sub 1(x)

(P/1-P) = theta sub 0 + theta sub 1 (x)
restriction in the range - restrict the no of data points

log of odds of P = log (P/1-P) = theta sub 0 + theta sub 1 (x)

exp[log (P/1-P) = exp [theta sub 0 + theta sub 1 (x)]

(p/1-p) = exp [theta sub 0 + theta sub 1 (x)]

P = exp(theta sub 0 + theta sub 1 (x)) - pexp(theta sub 0 + theta sub 1(x)

P = P [exp(theta sub 0 + theta sub 1 (x)/p - exp (theta sub 0 + theta sub 1 (x)]

(P/1-P) = theta sub 0 + theta sub 1 (x)
restriction in the range - restrict the no of data points

log of odds of P = log (P/1-P) = theta sub 0 + theta sub 1 (x)

exp[log (P/1-P) = exp [theta sub 0 + theta sub 1 (x)]

(p/1-p) = exp [theta sub 0 + theta sub 1 (x)]

P = exp(theta sub 0 + theta sub 1 (x)) - pexp(theta sub 0 + theta sub 1(x)

P = P [exp(theta sub 0 + theta sub 1 (x)/p - exp (theta sub 0 + theta sub 1 (x)]

P [1 + exp(theta sub 0 + theta sub 1(x)] = exp (theta sub 0 + theta sub 1 (x))

(P/1-P) = theta sub 0 + theta sub 1 (x)
restriction in the range - restrict the no of data points

log of odds of P = log (P/1-P) = theta sub 0 + theta sub 1 (x)

exp[log (P/1-P) = exp [theta sub 0 + theta sub 1 (x)]

(p/1-p) = exp [theta sub 0 + theta sub 1 (x)]

P = exp(theta sub 0 + theta sub 1 (x)) - pexp(theta sub 0 + theta sub 1(x)

P = P [exp(theta sub 0 + theta sub 1 (x)/p - exp (theta sub 0 + theta sub 1 (x)]

P [1 + exp(theta sub 0 + theta sub 1(x)] = exp (theta sub 0 + theta sub 1 (x))

H(x) = P = 1/(1+ exp^(-theta sub 0 - theta sub 1 (x))

$$\exp[\log(\frac{p}{1-p})] = \exp(\beta_0 + \beta_1 x)$$

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

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

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

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

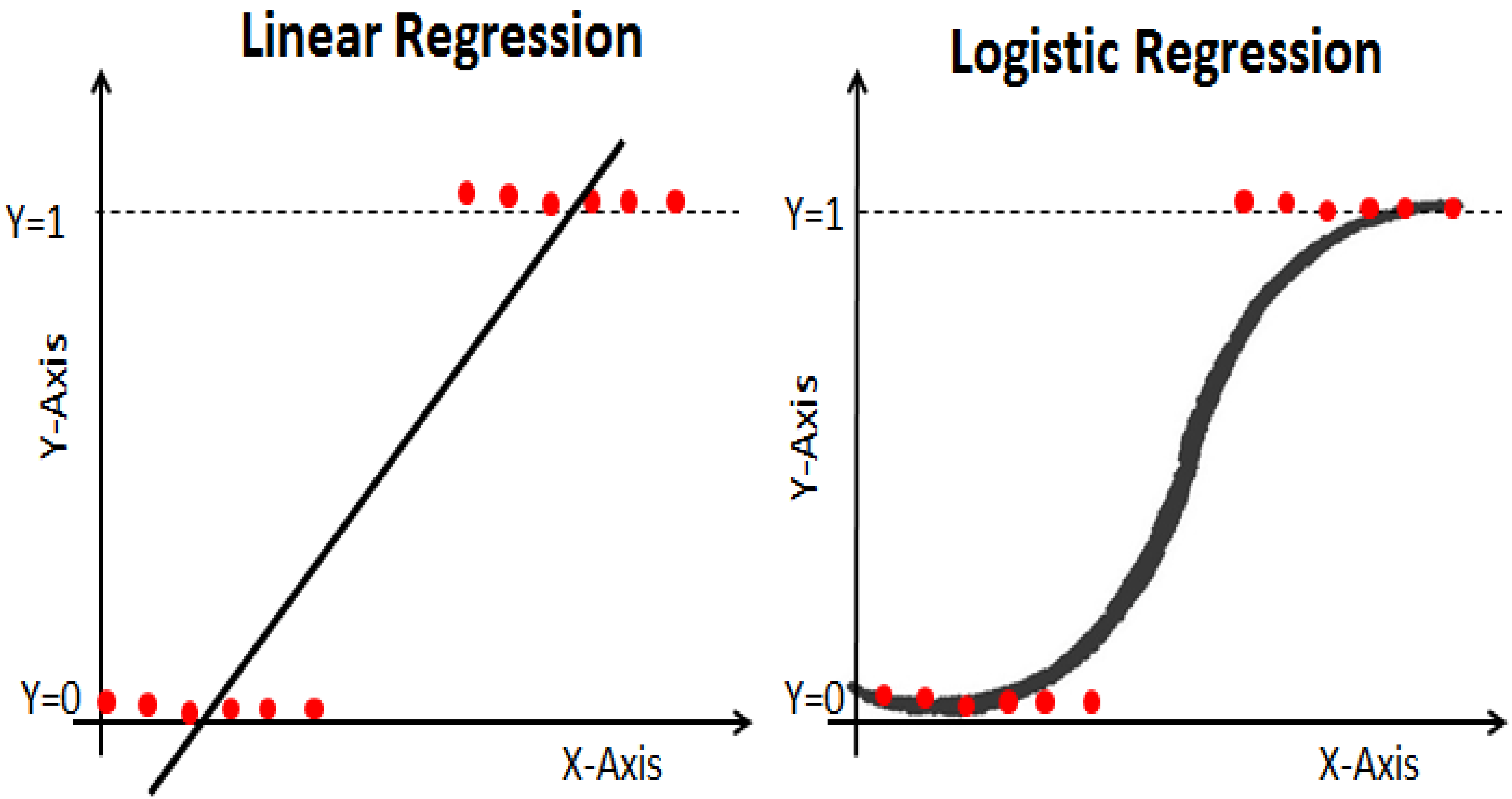
$$1 = \frac{e^{\left(\beta_0 + \beta_1 x\right)}}{p} - e^{\left(\beta_0 + \beta_1 x\right)}$$

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

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

Now dividing by $e^{\left(\beta_0 + \beta_1 x\right)}$*, we will get*

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$ *This is our sigmoid function.*



www.datacamp.com