Linear Regression

Print slope , intercept & line

Linear regression

In linear regression, the equation that describes the factor-response relationships is $Y=mX+C$ where Y and X are vectors that describe the response variable and the factor variable respectively.

Multiple regression

In a multiple regression case, we're interested in the impact of not only one, but many different factors, on the response variable.

The equation can be written as $Y=m_1X_1+m_2X_2+C$ where m1 and m2 are the coefficients of factors X1 and X2 respectively.

Polynomial Regression

For polynomial regression, we might use higher powers of X to describe Y, as described in

$Y=m_1X+m_2 X^2+C$

where m1 and m2 are coefficients of the first and second powers of the factor.

Logistic regression

Logistic regression is a statistical method for analysing a dataset in which there are one or more independent variables that determine an outcome.

The outcome is measured with a variable in which there are only two possible outcomes. It is used to predict a binary outcome (1 / 0, Yes / No, True / False) given a set of independent variables.

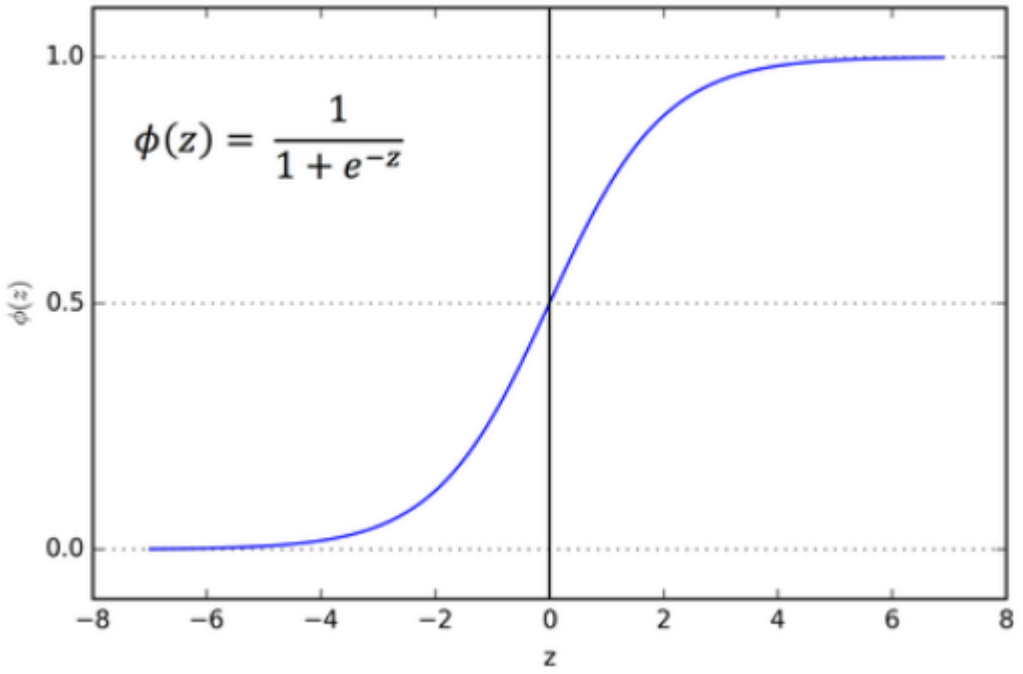
Applications

- Spam Detection : Predicting if an email is Spam or not
- Credit Card Fraud : Predicting if a given credit card transaction is fraud or not
- Health : Predicting if a given mass of tissue is benign or malignant
- Marketing : Predicting if a given user will buy an insurance product or not
- Banking : Predicting if a customer will default on a loan

sigmoid/logistic function

In logistic regression, the goal is to determine a mathematical equation that can be used to predict the probability of event

The sigmoid/logistic function is given by the following equation.



Metrics for Performance Evaluation

- Focus on the predictive capability of a model
  - Rather than how fast it takes to classify or build models, scalability, etc.
- Confusion Matrix:

		PREDICTED CLASS	
		Class=Yes	Class=No
ACTUAL CLASS	Class=Yes	a	b
	Class=No	c	d

a: TP (true positive)

b: FN (false negative)

c: FP (false positive)

d: TN (true negative)

Metrics for Performance Evaluation…

		PREDICTED CLASS	
		Class=Yes	Class=No
ACTUAL CLASS	Class=Yes	a (TP)	b (FN)
	Class=No	c (FP)	d (TN)

- Most widely-used metric:

$$\text{Accuracy} = \frac{a+d}{a+b+c+d} = \frac{TP+TN}{TP+TN+FP+FN}$$

Metrics for Performance Evaluation…

		PREDICTED CLASS	
		Class=Yes	Class=No
ACTUAL CLASS	Class=Yes	a (TP)	b (FN)
	Class=No	c (FP)	d (TN)

$$\text{precision} = \frac{tp}{tp + fp},$$

$$\text{recall} = \frac{tp}{tp + fn},$$

Precision and Recall

Precision is the fraction of information retrieved that are relevant.

Recall is the fraction of relevant information that are retrieved.

	PREDICTED CLASS		
ACTUAL CLASS		Class=Yes	Class=No
	Class=Yes	a (TP)	b (FN)
	Class=No	c (FP)	d (TN)

$$\text{precision} = \frac{tp}{tp + fp},$$

$$\text{recall} = \frac{tp}{tp + fn},$$

F1 Score

Harmonic mean of precision and recall

$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$