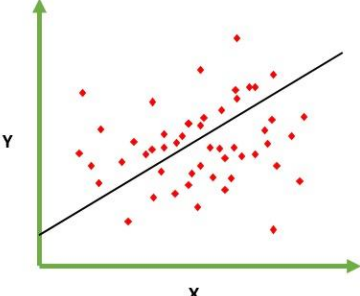
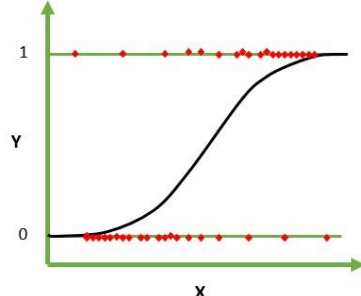


Difference between Linear Regression and Logistic Regression:

BASIS FOR COMPARISON	LINEAR REGRESSION	LOGISTIC REGRESSION
Basic	The data is modelled using a straight line.	The probability of some obtained event is represented as a linear function of a combination of predictor variables.
Linear relationship between dependent and independent variables	Is required	Not required
The independent variable	Could be correlated with each other. (Specially in multiple linear regression)	Should not be correlated with each other (no multicollinearity exist).
Example Diagram:	 A scatter plot with a vertical y-axis and a horizontal x-axis, both marked with green arrows. Red diamond-shaped data points are scattered across the plot, showing a positive linear trend. A solid black straight line is drawn through the data points, representing the linear regression model.	 A scatter plot with a vertical y-axis and a horizontal x-axis, both marked with green arrows. The y-axis has tick marks at 0 and 1. Red diamond-shaped data points are clustered at y=0 and y=1. A solid black S-shaped (sigmoid) curve is drawn through the data points, representing the logistic regression model.

Difference between Data Scientist and Data Engineer:

Data Scientist:

Data scientists aren't like every other scientist. Data scientists do not wear white coats or work in high tech labs full of science fiction movie equipment. They work in offices just like you and me. What differs them from most of us is that they are math experts. They use linear algebra and multivariable calculus to create new insight from existing data.

Here's an example:

An industrial company produces a lot of products that need to be tested before shipping.

Usually such tests take a lot of time because there are hundreds of things to be tested. All to make sure that your product is not broken. Wouldn't it be great to know early if a test fails ten steps down the line? If you knew that you could skip the other tests and just trash the product or repair it.

That's exactly where a data scientist can help you, big-time. This field is called predictive analytics and the technique of choice is machine learning. You feed an algorithm with measurement data. It generates a model and optimises it based on the data you fed it with. That model basically represents a pattern of how your data is looking. You show that model new data and the model will tell you if the data still represents the data you have trained it with.

Data Engineer

Data Engineers are the link between the management's big data strategy and the data scientists that need to work with data. What they do is building the platforms that enable data scientists to do their magic.

These platforms are usually used in five different ways:

- ❖ Data ingestion and storage of large amounts of data
- ❖ Algorithm creation by data scientists
- ❖ Automation of the data scientist's machine learning models and algorithms for production use
- ❖ Data visualisation for employees and customers
- ❖ Most of the time these guys start as traditional solution architects for systems that involve SQL databases, web servers, SAP installations and other "standard" systems.

But to create big data platforms the engineer needs to be an expert in specifying, setting up and maintaining big data technologies like: Hadoop, Spark, HBase, Cassandra, MongoDB, Kafka, Redis and more. What they also need is experience on how to deploy systems on cloud infrastructure like at Amazon or Google or on-premise hardware.