**QS) What is Machine Learning? Why is it important? Name ML Algorithms (Atleast 10).**

- Machine learning (ML) is a **type of artificial intelligence** (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so.  Machine learning algorithms use historical data as input to predict new output values.

-**Machine learning is important** because it gives enterprises a view of trends in customer behavior and business operational patterns, as well as supports the development of new products. Many of today's leading companies, such as Facebook, Google and Uber, make machine learning a central part of their operations. Machine learning has become a significant competitive differentiator for many companies.

- Machine learning algorithms are **classified into** 4 types:

Supervised and

Unsupervised Learning

Semi-supervised Learning

Reinforcement Learning

**#Top 10** **commonly used Machine Learning (ML) Algorithms**:

Linear regression

Logistic regression

Decision tree

SVM algorithm

Naive Bayes algorithm

KNN algorithm

K-means

Random forest algorithm

Dimensionality reduction algorithms

Gradient boosting algorithm and AdaBoosting algorithm



QS) What is **Linear Regression**? Explain the mathematics.

Linear Regression- process of estimating an unknown quantity based on some known ones (this is the regression part) with the condition that the unknown quantity can be obtained from the known ones by using only 2 operations: scalar multiplication and addition (this is the linear part). We multiply each known quantity by some number, and then we add all those terms to obtain an estimate of the unknown one.

Used in:

* **Evaluating tends and sales estimates**

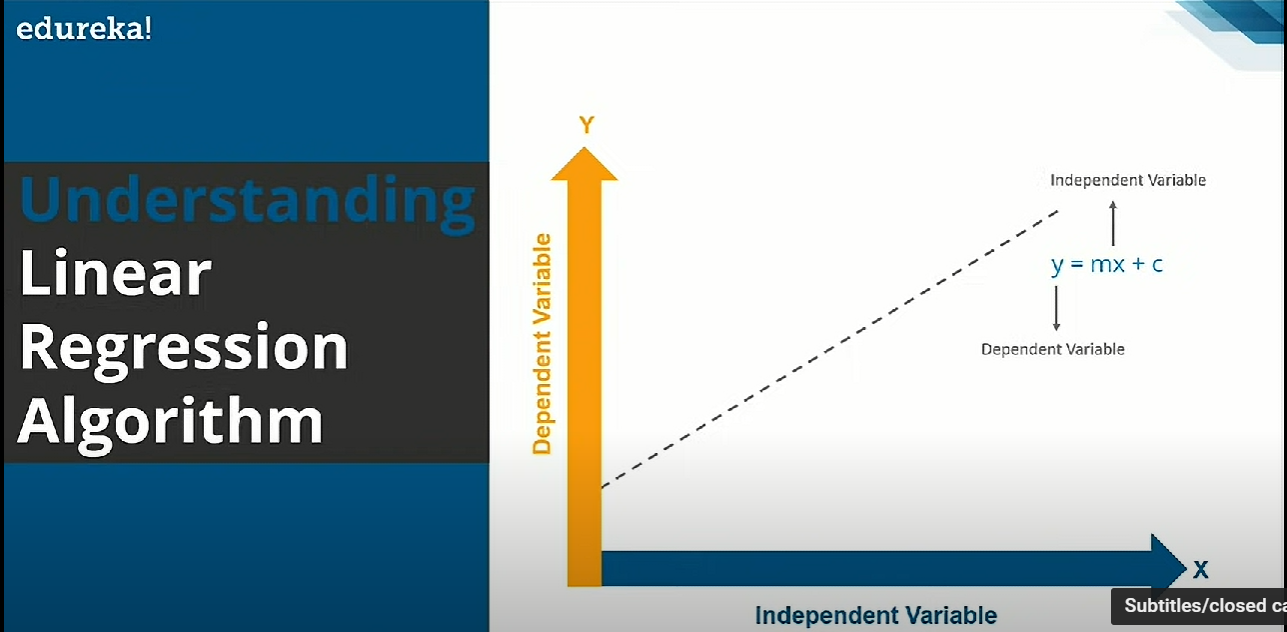
For example, if a company sales have increased steadily every month for past few years, then conducting a linear analysis on the sales data with monthly sales on the y axis and time on the x axis. This will give you a line that predicts the upward trends in the sale after creating the trend line the company could use the slope of the lines to focus sale in future months.

* **Analyzing the impact of price changes**

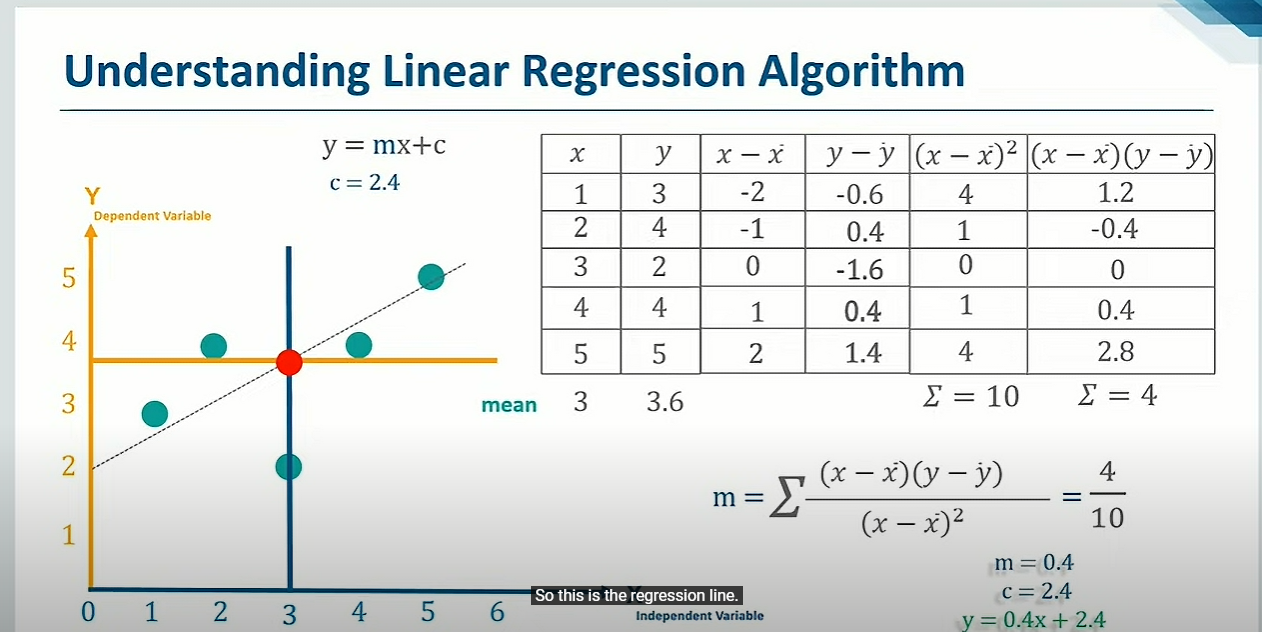
Linear regression can be used to analyze the effect of pricing on consumer behavior. For instance: If a company changes the price on a certain product several times, then it can record the quantity itself for each price level and then perform a linear regression with sold quantity as a dependent variable and price as the independent variable. This would result in a line that depicts the extent to which they reduce their consumption of the product as the price is increasing. So this result would help us in future pricing decisions.

* **Assessment of risk in financial services and insurance domain:**

Linear regression can be used to analyze the risk, for example health insurance company might conduct a linear regression algorithm how it can do it can do it by plotting the number of claims per customer against its age and they might discover that the old customers then to make more health insurance claim. Well the result of such analysis might guide important business decisions.

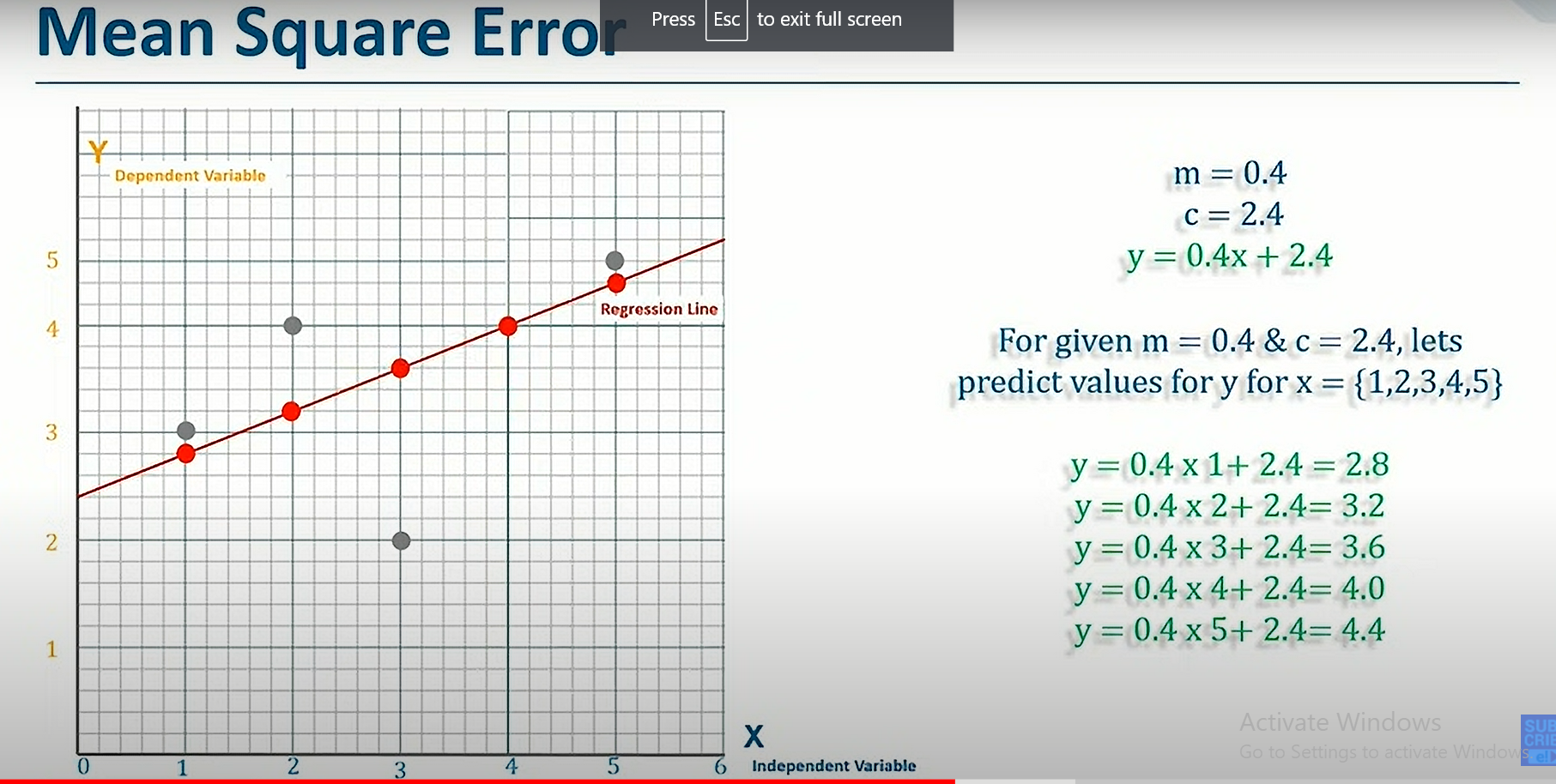


The math behind it:



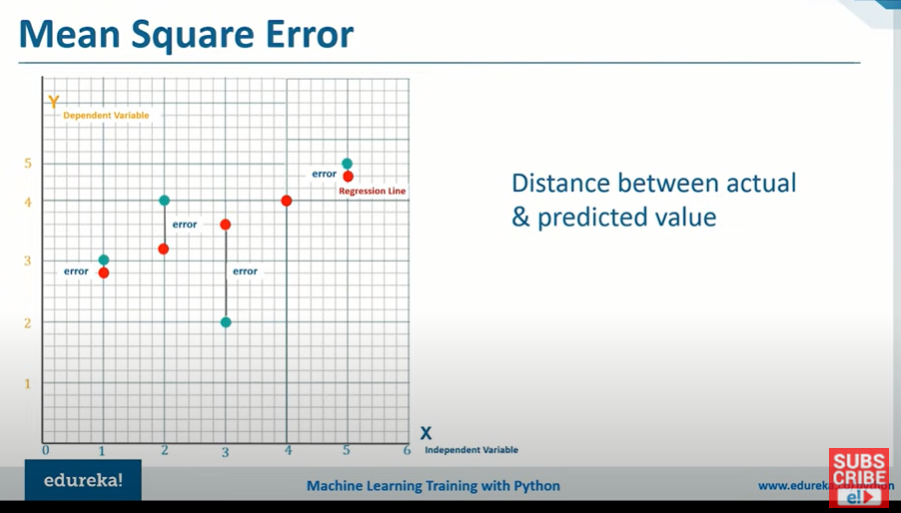
^Red dot= mean of x, mean of y = (3, 3.6).

The green dots are all values of (x, y) plotted.



^^Regression line:

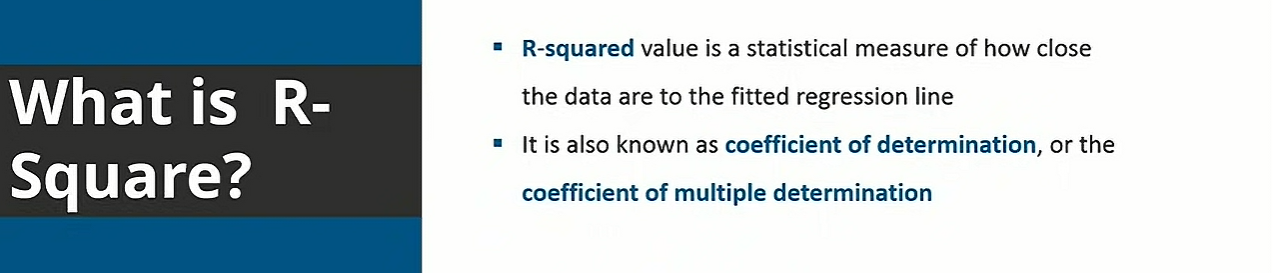
The line passing through all these predicted values and cutting y axis at 2.4.

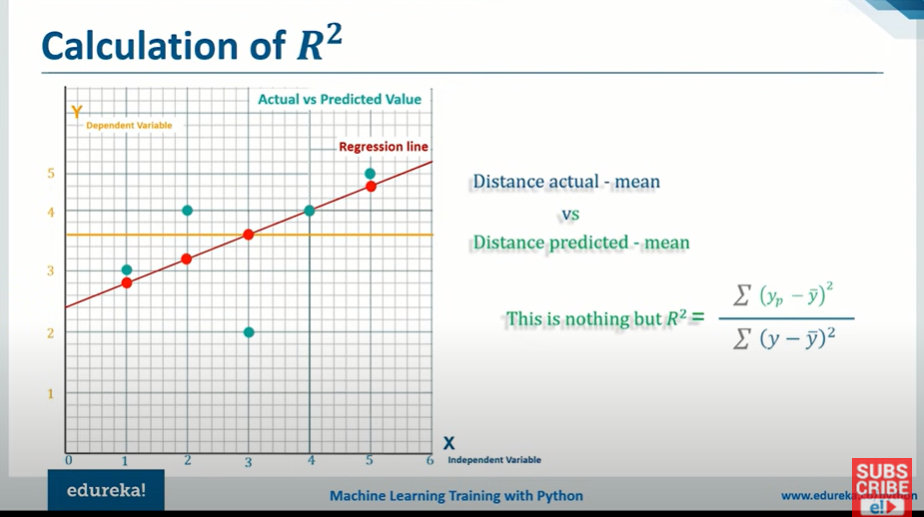


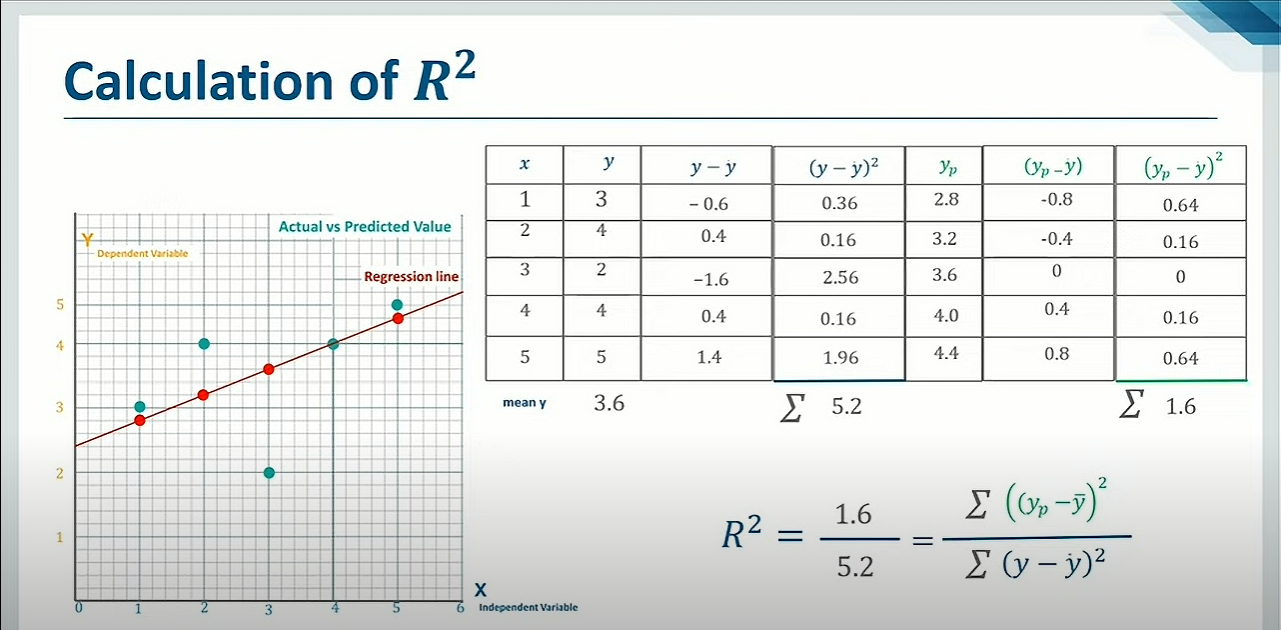
^ Line with least error = regression line and also the best fit line.

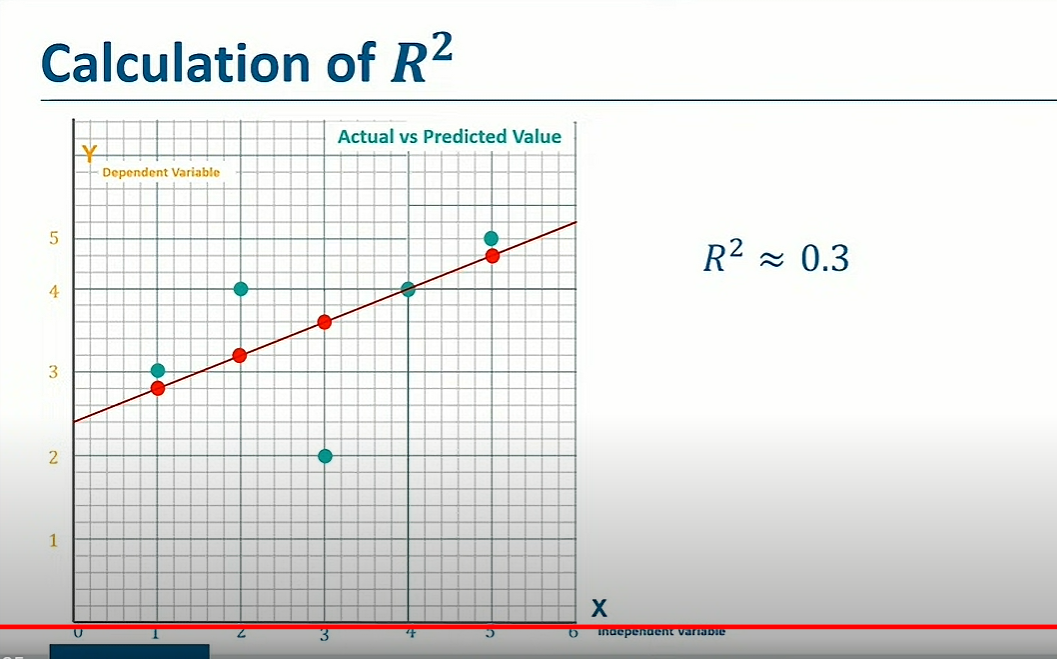
--------

Checking the Goodness of fit:









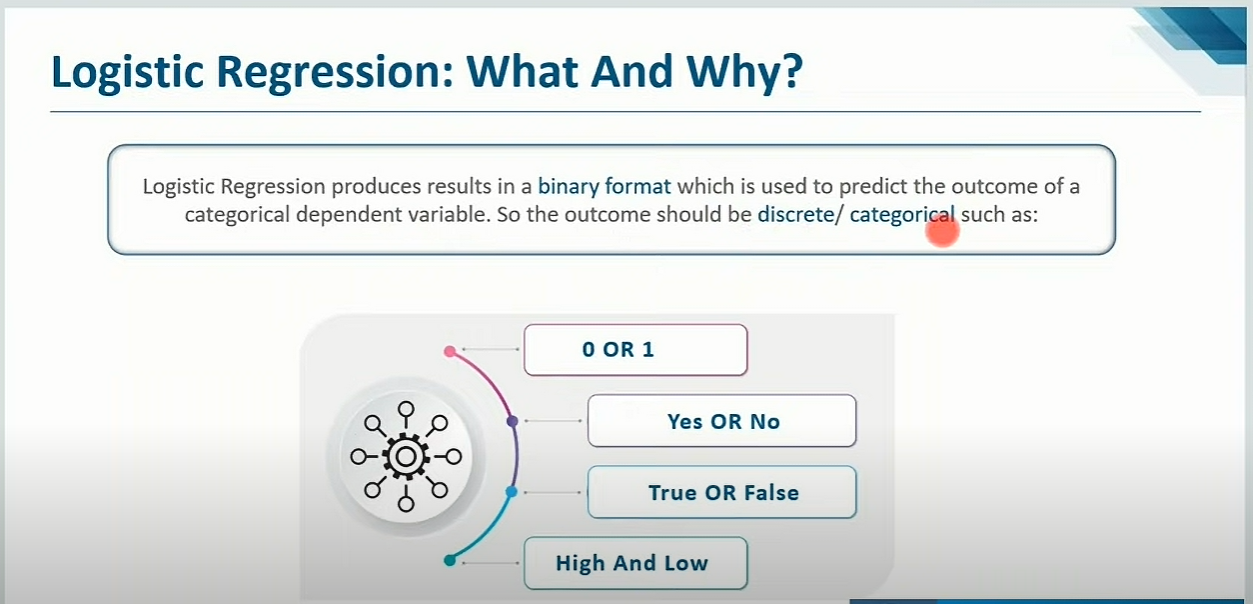
This is shows the fit isn't good as the data points are far away from the regression line.

If R^2 = 1:



The actual values lie on the regression line itself.

QS) What is **Logistic Regression**? Explain the mathematics.



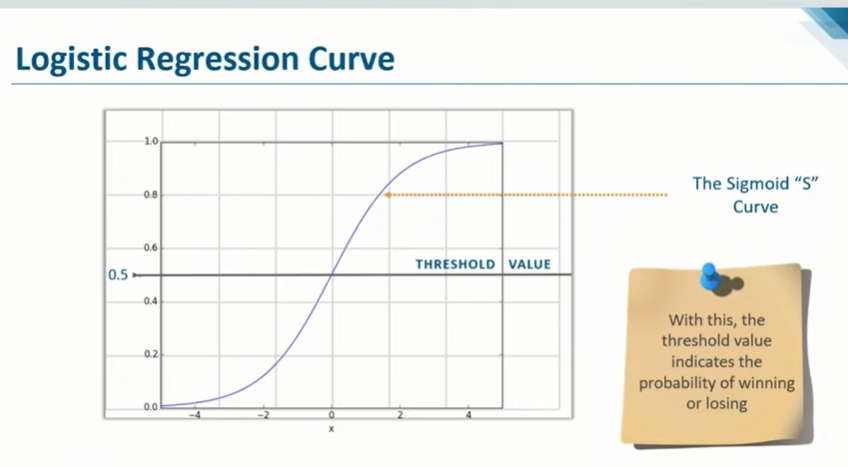
In linear regression, there are many values in the range 0-1. However, in the logistical regression the values can only be 0 or 1 in the range 0-1.

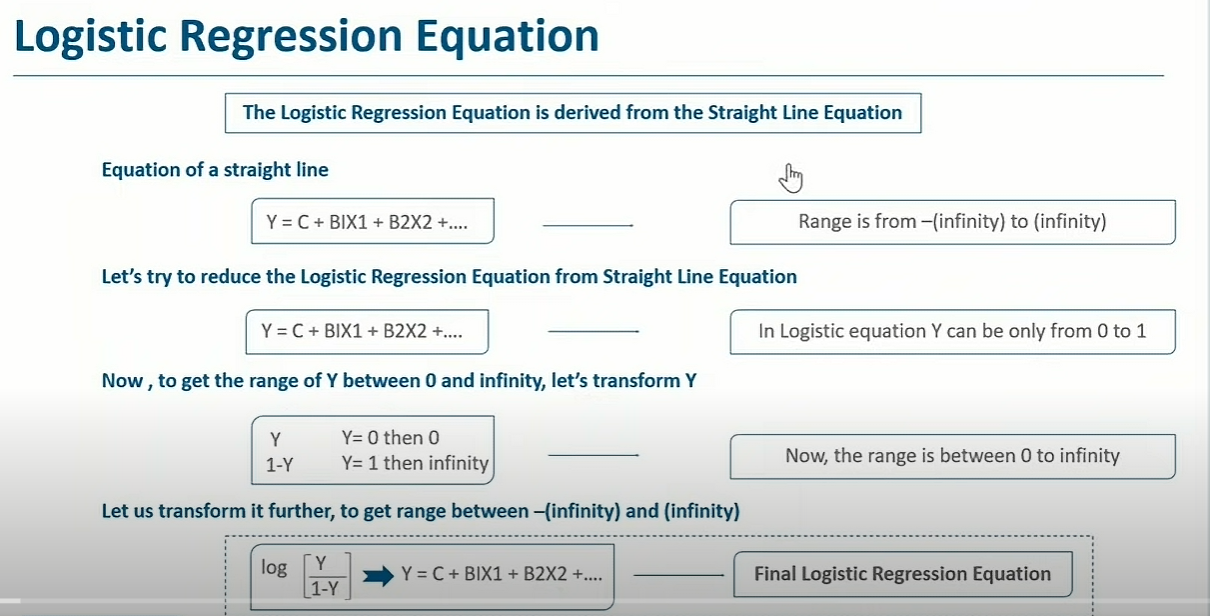
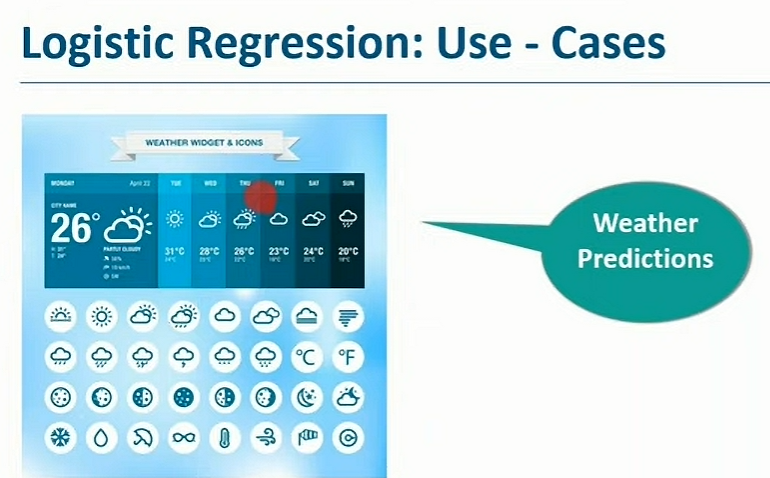
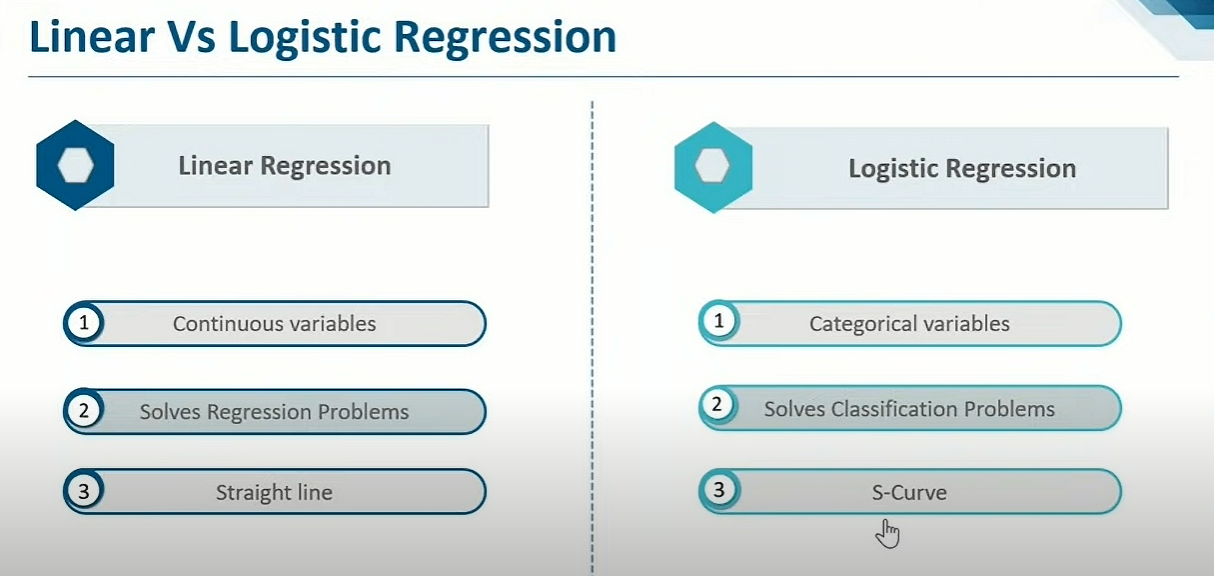
Sigmoid "S" curve- converts any values from - infinity to infinity to discrete values (aka 0 or 1)

The sigmoid function has values very close to either 0 or 1 across most of its domain.

If we have a data point 0.8, how do we determine if it's 0 or 1?

For this we have threshold value. Winning (more than threshold value) = 1 , losing (less than threshold value)= 0.



Logistic regression: whether it rains or not?

(while linear regression would be what is the temp tomorrow? )

