# Linear Regression

- Logistic Regression is actually a classification technique.
- Linear Regression is actual Regression technique.

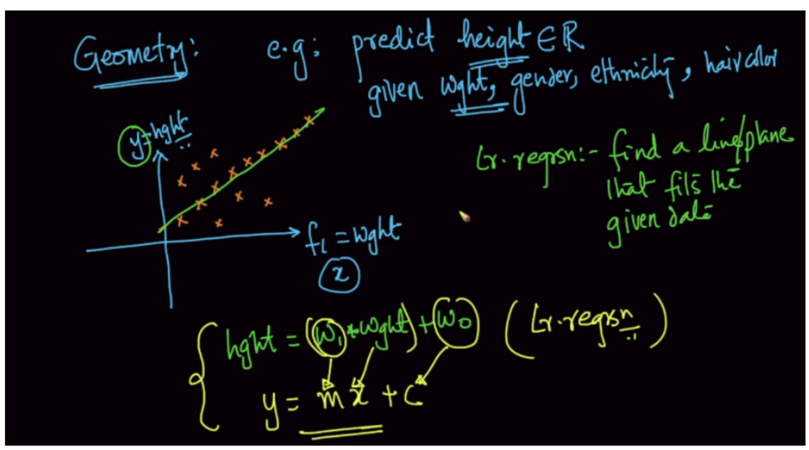
**Correction**- Real World Cases- Mistake at 00:12, there is no problem of imbalanced data in linear regression as there are no class labels

**Linear regression** is a linear model, i.e. a model that assumes a linear relationship between the input variables (x) and the single output variable (y).

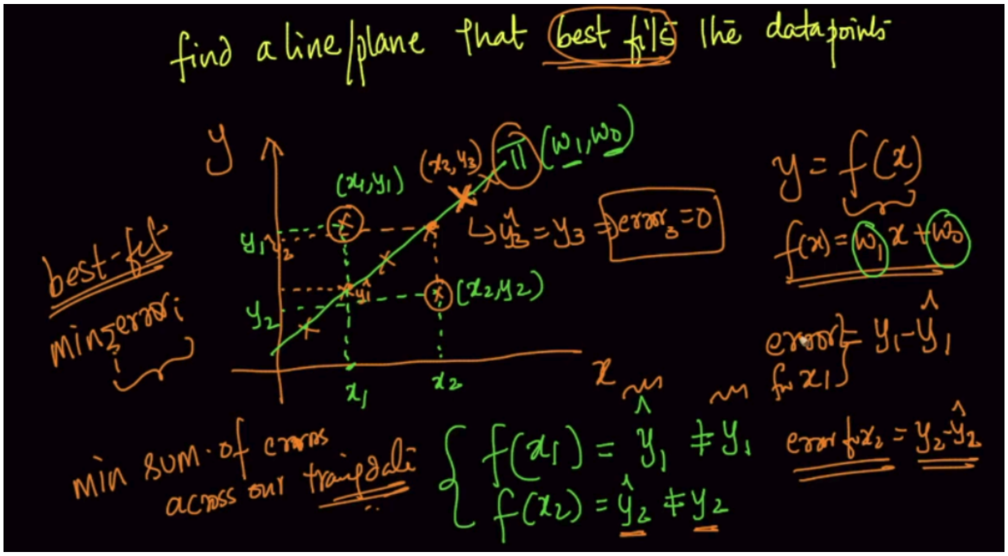
- More specifically, that y can be calculated from a linear combination of the input variables (x).
- When there is a single input variable (x), the method is referred to as simple linear regression.
- When there are multiple input variables, then it refers to the method as multiple linear regression.

Different techniques can be used to prepare or train the linear regression equation from data, the most common of which is called Ordinary Least Squares and Gradient Descent.

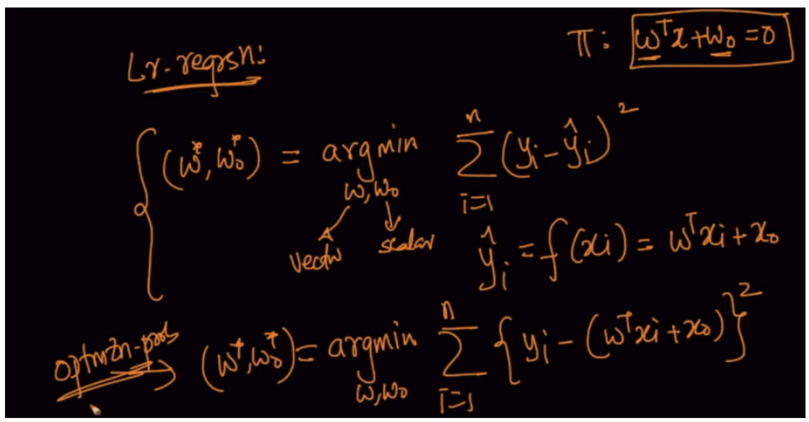
Refer- [https://machinelearningmastery.com/linear-regression-for-machine-learning/#:~:text=Linear%20regression%20is%20a%20linear,single%20output%20variable%20(y).&te (https://machinelearningmastery.com/linear-regression-for-machine-learning/#:~:text=Linear%20regression%20is%20a%20linear,single%20output%20variable%20(y).&te]



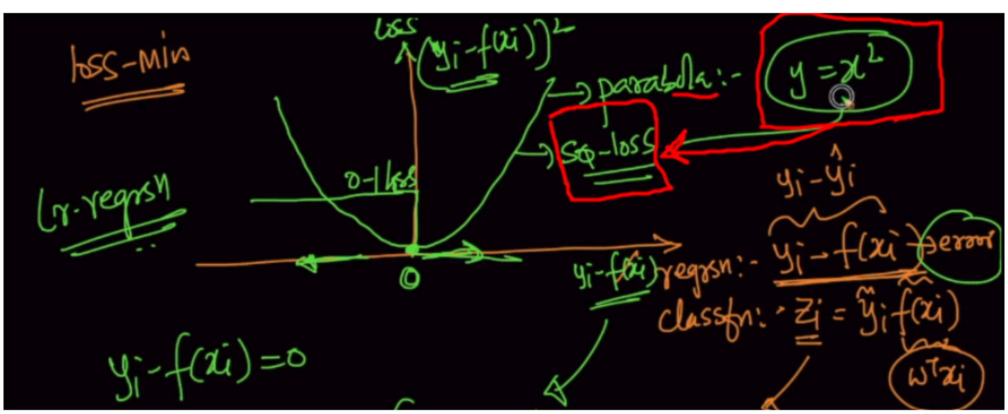

- In linear regression, the algorithm tries to best fit the points with a linear surface having intercept w_0.
- There will be scenarios where many a point will not fall on the linear surface.
- To resolve this issue, we project those points on the linear surface with y transformed into ŷ. Therefore, the error for the particular points would be = (y - ŷ)




## The mathematical formulation of Linear Regression



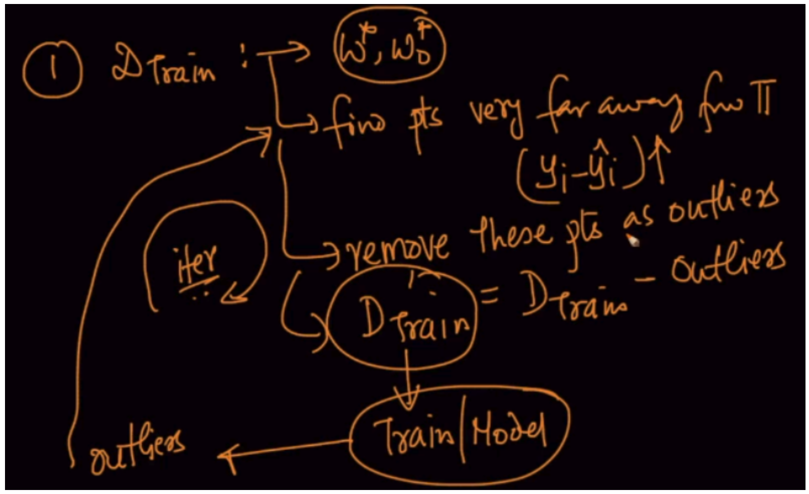

We need to regularize the equation by L1 or L2 regularization to avoid underfitting or overfitting with lambda as the hyperparameter




## Real world cases -

1. There is no problem with imbalanced data as there are no class labels.
2. Features should not be multicollinear.
3. L1 regularization creates sparsity.

4. Since Linear Regression can be heavily impacted by outliers, follow the technique shown below.




Refer-

1. https://towardsdatascience.com/linear-regression-in-real-life-4a78d7159f16 (https://towardsdatascience.com/linear-regression-in-real-life-4a78d7159f16)
2. https://www.kaggle.com/cengizeralp/practice-2-suicide-analysis-with-regression (https://www.kaggle.com/cengizeralp/practice-2-suicide-analysis-with-regression)

1. Linear regression with L2 regularization is also referred to as "Ridge Regression"
2. Linear regression with L1 regularization is also referred to as "Lasso Regression".
3. Linear regression with l1+L2 regularization is also referred to as "ElasticNet Regression".

Refer -
Need for regularization in Linear Regression -
https://www.youtube.com/watchv=iUm2Z1SKuGk&feature=youtu.be
(https://www.youtube.com/watchv=iUm2Z1SKuGk&feature=youtu.be)

## Linear Regression Learning the Model -

Learning a linear regression model means estimating the values of the coefficients used in the representation with the data that we have available.

In this section we will take a brief look at four techniques to prepare a linear regression model.

1. Simple Linear Regression
2. Ordinary Least Squares
3. Gradient Descent
4. Regularization

**1. Simple Linear Regression -**

With simple linear regression when we have a single input, we can use statistics to estimate the coefficients.

This requires that you calculate statistical properties from the data such as means, standard deviations, correlations and covariance. All of the data must be available to traverse and calculate statistics.

**2. Ordinary Least Squares -**

When we have more than one input we can use Ordinary Least Squares to estimate the values of the coefficients.

The Ordinary Least Squares procedure seeks to minimize the sum of the squared residuals. This means that given a regression line through the data we calculate the distance from each data point to the regression line, square it, and sum all of the squared errors together. This is the quantity that ordinary least squares seeks to minimize.

This approach treats the data as a matrix and uses linear algebra operations to estimate the optimal values for the coefficients. It means that all of the data must be available and you must have enough memory to fit the data and perform matrix operations.

**3. Gradient Descent**

When there are one or more inputs we can use a process of optimizing the values of the coefficients by iteratively minimizing the error of the model on your training data.

This operation is called Gradient Descent and works by starting with random values for each coefficient. The sum of the squared errors are calculated for each pair of input and output values. A learning rate is used as a scale factor and the coefficients are updated in the direction towards minimizing the error. The process is repeated until a minimum sum squared error is achieved or no further improvement is possible.

In practice, it is useful when you have a very large dataset either in the number of rows or the number of columns that may not fit into memory.

**4. Regularization**

It seek to both minimize the sum of the squared error of the model on the training data (using ordinary least squares) but also to reduce the complexity of the model.

Two popular examples of regularization procedures for linear regression are:

- **Lasso Regression** where Ordinary Least Squares is modified to also minimize the absolute sum of the coefficients (called L1 regularization).
- **Ridge Regression** where Ordinary Least Squares is modified to also minimize the squared absolute sum of the coefficients (called L2 regularization).

These methods are effective to use when there is collinearity in your input values and ordinary least squares would overfit the training data.