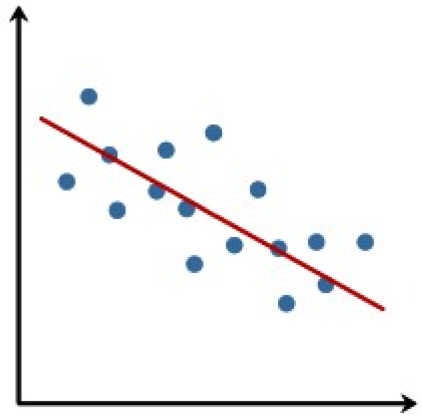
Linear Regression



Tushar B. Kute,
http://tusharkute.com

Linear Regression

- Linear regression is used for finding linear relationship between target and one or more predictors.
- There are two types of linear regression-
  - Simple and
  - Multiple.

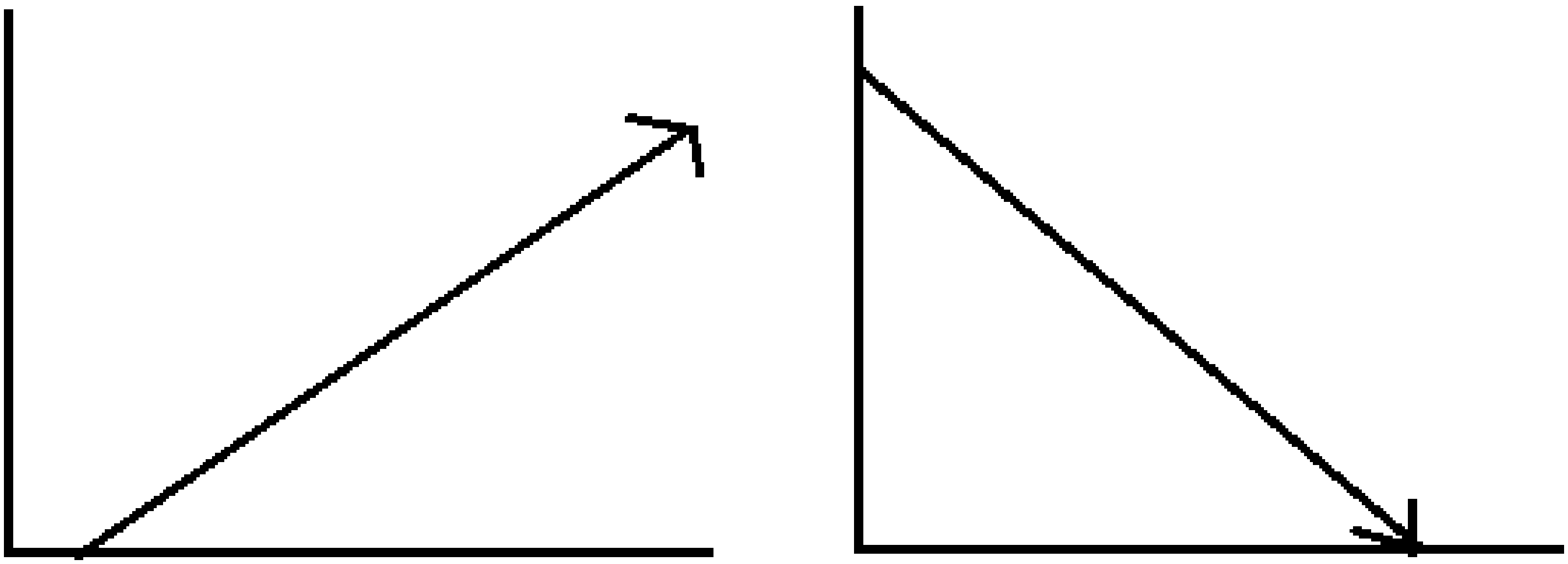
Simple Linear Regression

- Simple linear regression is useful for finding relationship between two continuous variables. One is predictor or independent variable and other is response or dependent variable.
- It looks for statistical relationship but not deterministic relationship. Relationship between two variables is said to be deterministic if one variable can be accurately expressed by the other.
- For example, using temperature in degree Celsius it is possible to accurately predict Fahrenheit.
- Statistical relationship is not accurate in determining relationship between two variables. For example, relationship between height and weight.

Core Idea

- The core idea is to obtain a line that best fits the data.
- The best fit line is the one for which total prediction error (all data points) are as small as possible.
- Error is the distance between the point to the regression line.

Linear Regression



Positive Linear Relationship

Negative Linear Relationship

Real Life Example

- We have a dataset which contains information about relationship between ‘number of hours studied’ and ‘marks obtained’.
- Many students have been observed and their hours of study and grade are recorded. This will be our training data.
- Goal is to design a model that can predict marks if given the number of hours studied. Using the training data, a regression line is obtained which will give minimum error.
- This linear equation is then used for any new data.

Real Life Example

- if we give number of hours studied by a student as an input, our model should predict their mark with minimum error.

  Y(pred) = b0 + b1*x

- The values b0 and b1 must be chosen so that they minimize the error. If sum of squared error is taken as a metric to evaluate the model, then goal to obtain a line that best reduces the error.

$$\text{Error} = \sum_{i=1}^{n}(actual\_output - predicted\_output) ** 2$$

- If we don’t square the error, then positive and negative point will cancel out each other.

Many names of Linear Regression

- The reason is because linear regression has been around for so long (more than 200 years). It has been studied from every possible angle and often each angle has a new and different name.
- Linear regression is a linear model, e.g. 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, literature from statistics often refers to the method as multiple linear regression.

Applications

- **Trend lines**: A trend line represents the variation in some quantitative data with passage of time (like GDP, oil prices, etc.). These trends usually follow a linear relationship. Hence, linear regression can be applied to predict future values.
- **Economics**: To predict consumption spending, fixed investment spending, inventory investment, purchases of a country's exports, spending on imports, the demand to hold liquid assets, labor demand, and labor supply.
- **Finance**: Capital price asset model uses linear regression to analyze and quantify the systematic risks of an investment.
- **Biology**: Linear regression is used to model causal relationships between parameters in biological systems.

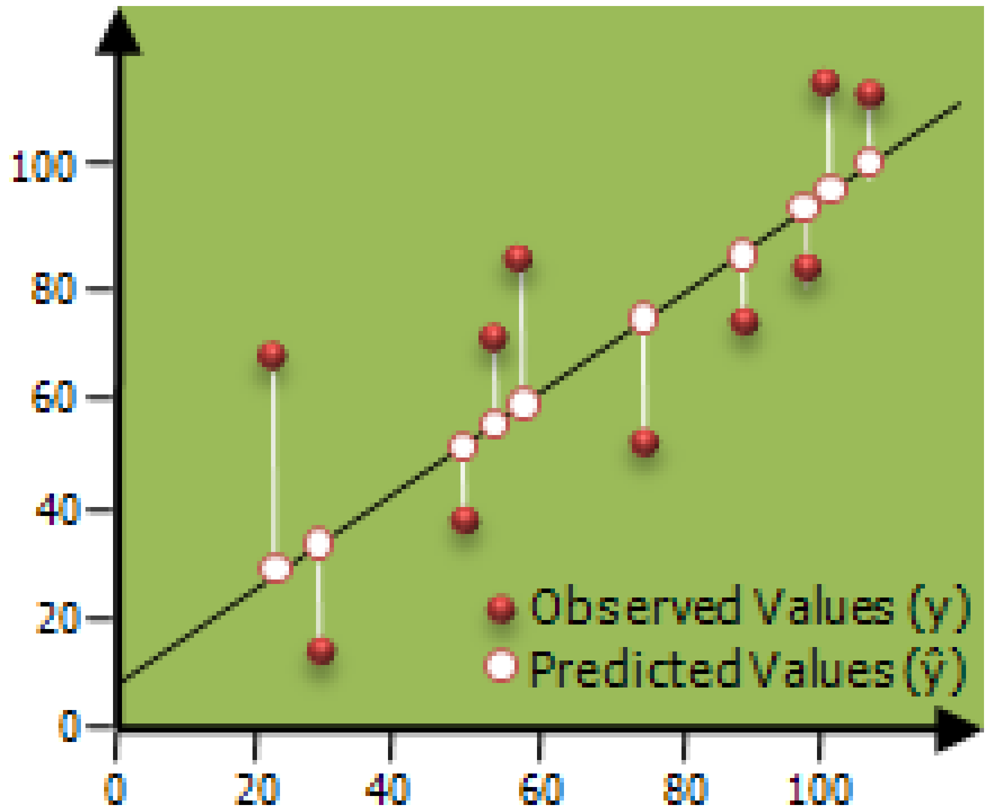
Ordinary Least Square

- 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.

Ordinary Least Square

- 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.
- It is unusual to implement the Ordinary Least Squares procedure yourself unless as an exercise in linear algebra.
- It is more likely that you will call a procedure in a linear algebra library. This procedure is very fast to calculate.

Ordinary Least Square



Gradient Descent

- When there are one or more inputs you 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.

Regularization

- There are extensions of the training of the linear model called regularization methods. These 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 (like the number or absolute size of the sum of all coefficients in 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).

Example:

- Go practical...

Thank you

@mitu_skillologies

/mITuSkillologies

@mitu_group

/company/mitu-skillologies

MITUSkillologies

Web Resources
https://mitu.co.in
http://tusharkute.com

contact@mitu.co.in

tushar@tusharkute.com