**Explain simple linear regression.**

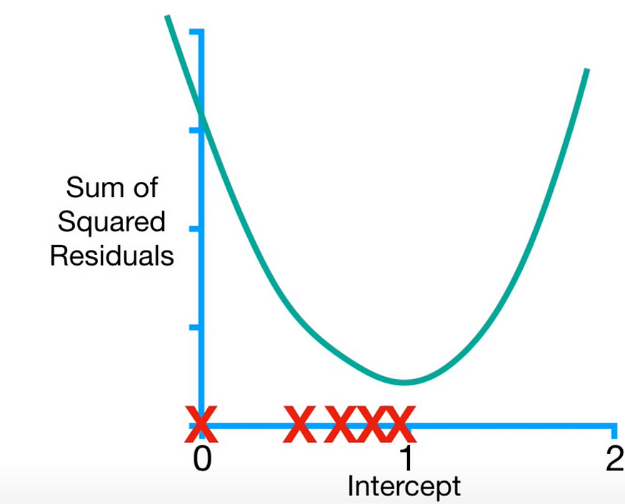
* Simple linear regression is a statistical method for establishing the relationship between two variables using a straight line.
* Simple linear regression is a statistical method that allows us to summarize and study relationships between two continuous (quantitative) variables:
* One variable, denoted x, is regarded as the predictor, explanatory, or independent variable.
* The other variable, denoted y, is regarded as the response, outcome, or dependent variable.
* The goal of simple linear regression is to find the best-fitting straight line that describes the relationship between the independent variable (often denoted as *x*) and the dependent variable (often denoted as *y*).

**y = β0 +β1x+ε** is the formula used for simple linear regression.

* y is the predicted value of the dependent variable (y) for any given value of the independent variable (x).
* B0 is the intercept, the predicted value of y when the x is 0.
* B1 is the regression coefficient – how much we expect y to change as x increases.
* x is the independent variable ( the variable we expect is influencing y).
* e is the error of the estimate, or how much variation there is in our regression coefficient estimate.
* Linear regression finds the line of best fit line through your data by searching for the regression coefficient (B1) that minimizes the total error (e) of the model.
* The goal of simple linear regression is to estimate the values of *β*0​ and *β*1​ that minimize the sum of the squared differences between the observed values of *y* and the values predicted by the regression line. This process is often done using the method of least squares.
* Once the regression coefficients (*β*0​ and *β*1​) are estimated, the regression line can be plotted, allowing for predictions of the dependent variable (*y*) for any given value of the independent variable (*x*).

**Explain gradient descent for simple linear regression.**

* Gradient descent is one of the most famous techniques in machine learning and used for training all sorts of neural networks.
* Gradient Descent is a fundamental optimization algorithm used in machine learning and statistics to minimize a cost function.
* The main aim of gradient descent is to find the best parameters of a model which gives the highest accuracy on training as well as [testing datasets](https://www.geeksforgeeks.org/how-to-split-a-dataset-into-train-and-test-sets-using-python/).
* This makes Gradient descent very useful when it is not possible solve for where the derivate is equal to zero.



**Steps Required in Gradient Descent Algorithm**

* **Step 1** we first initialize the parameters of the model randomly
* **Step 2** Compute the gradient of the cost function with respect to each parameter. It involves making partial differentiation of cost function with respect to the parameters.
* **Step 3**Update the parameters of the model by taking steps in the opposite direction of the model. Here we choose a [hyperparameter learning rate](https://www.geeksforgeeks.org/hyperparameter-tuning/) which is denoted by alpha. It helps in deciding the step size of the gradient.
* **Step 4** Repeat steps 2 and 3 iteratively to get the best parameter for the defined model

**Advantages Of Gradient Descent**

* **Flexibility:** Gradient Descent can be used with various cost functions and can handle non-linear regression problems.
* **Scalability:** Gradient Descent is scalable to large datasets since it updates the parameters for each training example one at a time.
* **Convergence:** Gradient Descent can converge to the global minimum of the cost function, provided that the learning rate is set appropriately.

**Disadvantages Of Gradient Descent**

**Sensitivity to Learning Rate:** The choice of learning rate can be critical in Gradient Descent since using a high learning rate can cause the algorithm to overshoot the minimum, while a low learning rate can make the algorithm converge slowly.

**Slow Convergence:** Gradient Descent may require more iterations to converge to the minimum since it updates the parameters for each training example one at a time.

**Local Minima**: Gradient Descent can get stuck in local minima if the cost function has multiple local minima.

**What is hypothesis function for simple linear regression?**

* A hypothesis is an assumption that is made based on some evidence. This is the initial point of any investigation that translates the research questions into predictions.
* In the realm of machine learning, a hypothesis serves as an initial assumption made by data scientists and ML professionals when attempting to address a problem.
* Machine learning involves conducting experiments based on past experiences, and these hypotheses are crucial in formulating potential sol **Here are some types of hypotheses commonly used:**

**Null Hypothesis:**

* The null hypothesis suggests that there's no relationship between the independent and dependent variables.
* It assumes that the independent variable doesn't affect the dependent variable.
* Null hypotheses are often tested using statistical tests like t-tests and ANOVA.

**Alternative Hypothesis:**

* The alternative hypothesis contradicts the null hypothesis.
* It assumes that there is a relationship between the independent and dependent variables.
* Also known as a research hypothesis.

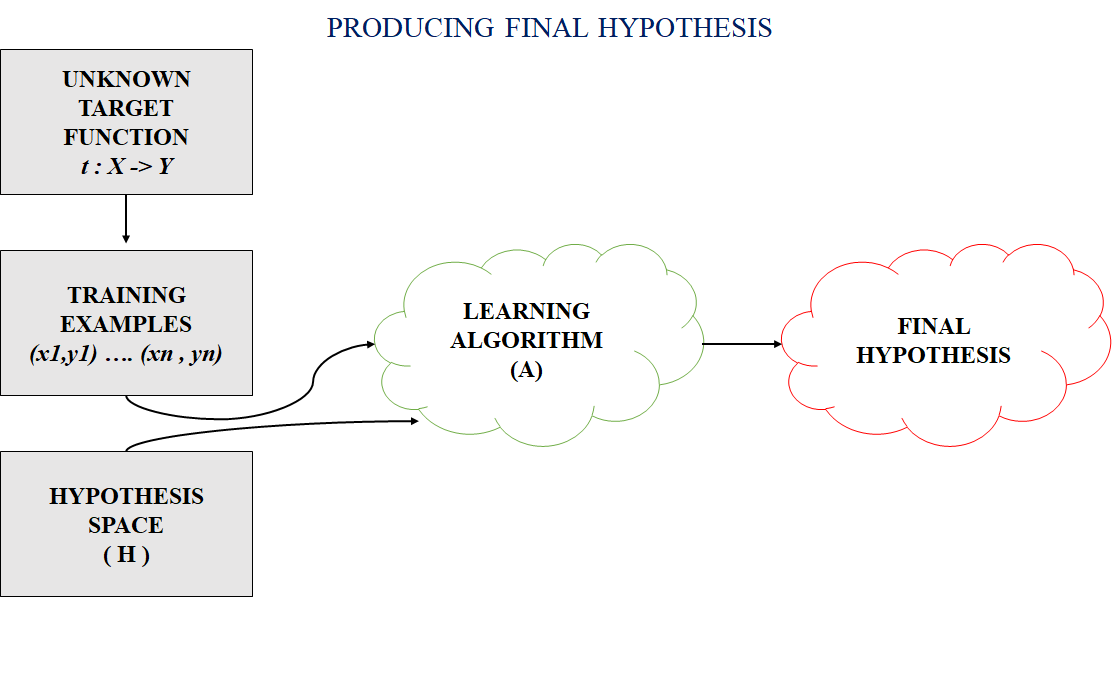
**One-tailed Hypothesis:**

* In a one-tailed test, the rejection region is located at one end of the sample distribution.
* It suggests that the estimated parameter is either more or less than the critical value, supporting the alternative hypothesis.
* Commonly used in chi-square distribution tests, it can be left-tailed or right-tailed.

**Two-tailed Hypothesis**:

* In a two-tailed test, the rejection region is on both ends of the normal distribution.
* It checks whether the sample falls within or outside a certain range of values.
* An alternative hypothesis is accepted if the calculated value falls in either tail of the probability distribution.

**Working of Hypothesis :**

****

The Hypothesis can be calculated as:

y = mx+b

Where,

* y = range
* m = slope of the lines
* x = domain
* b = intercept

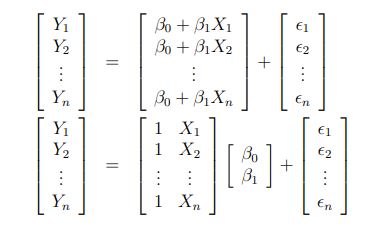
**Explain simple regression in matrix form.**

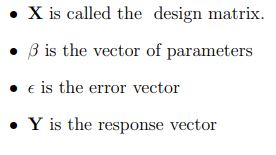
* Simple linear regression is a statistical method for establishing the relationship between two variables using a straight line.
* Simple linear regression is a statistical method that allows us to summarize and study relationships between two continuous (quantitative) variables:
* One variable, denoted x, is regarded as the predictor, explanatory, or independent variable.
* The other variable, denoted y, is regarded as the response, outcome, or dependent variable.
* The goal of simple linear regression is to find the best-fitting straight line that describes the relationship between the independent variable (often denoted as *x*) and the dependent variable (often denoted as *y*).

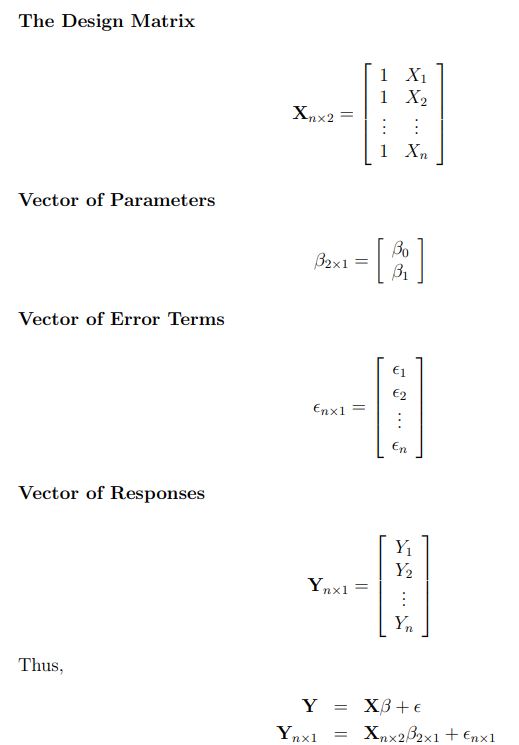
**y = β0 +β1x+ε** is the formula used for simple linear regression.

* y is the predicted value of the dependent variable (y) for any given value of the independent variable (x).
* B0 is the intercept, the predicted value of y when the x is 0.
* B1 is the regression coefficient – how much we expect y to change as x increases.
* x is the independent variable ( the variable we expect is influencing y).
* e is the error of the estimate, or how much variation there is in our regression coefficient estimate.

The Simple Linear Regression Model in Matrix Form :







**Explain Sampling Distribution of Estimators**

In machine learning, the concept of "sampling distribution of an estimator" is closely related to the field of statistics.

**Estimator:**

* An estimator is a statistical method or formula used to estimate a population parameter (such as mean, variance, etc.) based on a sample from that population.
* In machine learning, many algorithms act as estimators, as they try to learn patterns and relationships from a training dataset to make predictions on new, unseen data.

**Population and Sample:**

* In statistics, the population refers to the entire set of individuals, items, or data points that one is interested in studying.
* However, it is often impractical or impossible to gather data from the entire population.
* Hence, a sample is taken, which is a subset of the population, to draw conclusions about the population.

**Sampling Distribution:**

* The sampling distribution of an estimator is a theoretical probability distribution.
* It illustrates the range of potential values the estimator may assume when computed from various random samples of identical size drawn from the population.
* It describes the variability of the estimator's values and how it behaves across different samples.
* The sampling distribution of an estimator is the theoretical probability distribution that shows the possible values the estimator can take when calculated from different random samples of the same size from the population.
* In other words, if we were to repeatedly take samples from the same population and calculate the estimator for each sample, the sampling distribution would describe the distribution of those estimator values.
* Understanding the sampling distribution is crucial because it allows us to make statistical inferences and draw conclusions about the population parameter based on the behavior of the estimator.