Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed. Supervised learning and unsupervised learning are two main types of[machine learning](https://www.geeksforgeeks.org/machine-learning/).

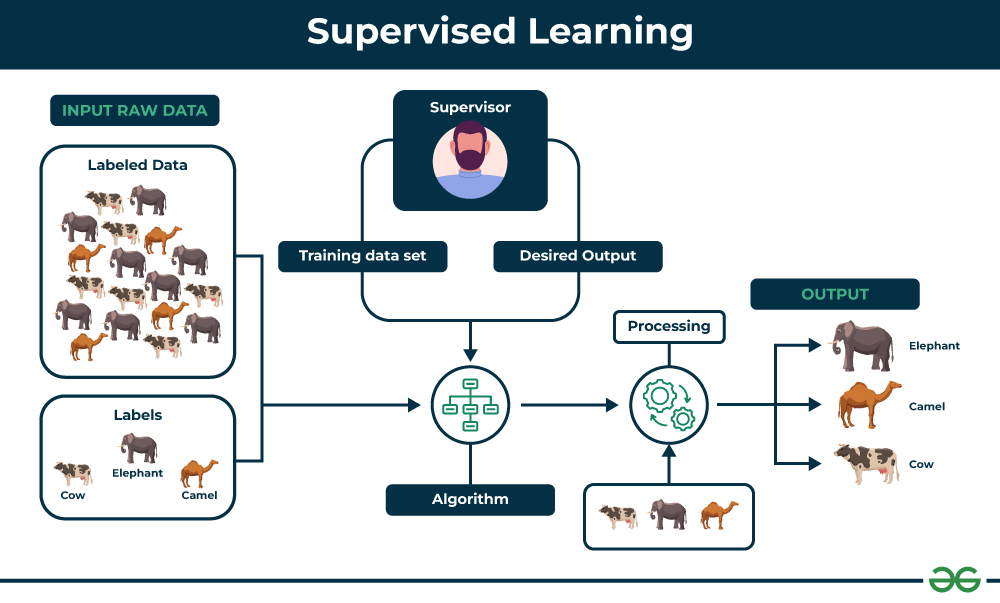
In [supervised learning](https://www.geeksforgeeks.org/supervised-machine-learning/), the machine is trained on a set of labeled data, which means that the input data is paired with the desired output.

In unsupervised learning, the machine is trained on a set of unlabeled data, which means that the input data is not paired with the desired output.

**What is Supervised learning?**

Supervised learning is a type of [machine learning algorithm](https://www.geeksforgeeks.org/machine-learning-algorithms/) that learns from labeled data. Labeled data is data that has been tagged with a correct answer or classification.

For example, a labeled dataset of images of Elephant, Camel and Cow would have each image tagged with either “Elephant” , “Camel”or “Cow.”



**Types of Supervised Learning**

Supervised learning is classified into two categories of algorithms:

* [**Regression**:](https://www.geeksforgeeks.org/regression-classification-supervised-machine-learning/) A regression problem is when the output variable is a real value, such as “dollars” or “weight”.
* [**Classification**:](https://www.geeksforgeeks.org/getting-started-with-classification/) A classification problem is when the output variable is a category, such as “Red” or “blue” , “disease” or “no disease”.

1- Regression

Regression is a type of supervised learning that is used to predict continuous values, such as house prices, stock prices, or customer churn. Regression algorithms learn a function that maps from the input features to the output value.

Some common [regression algorithms](https://www.geeksforgeeks.org/types-of-regression-techniques/) include:

* Linear Regression
* Polynomial Regression
* Support Vector Machine Regression
* Decision Tree Regression
* Random Forest Regression

**2- Classification**

Classification is a type of supervised learning that is used to predict categorical values, such as whether a customer will churn or not, whether an email is spam or not, or whether a medical image shows a tumor or not. Classification algorithms learn a function that maps from the input features to a probability distribution over the output classes.

Some common[classification algorithms](https://www.geeksforgeeks.org/top-6-machine-learning-algorithms-for-classification/) include:

* Logistic Regression
* Support Vector Machines
* Decision Trees
* Random Forests
* Naive Baye

**Evaluating Supervised Learning Models**

**For Regression**

* **Mean Squared Error (MSE):** MSE measures the average squared difference between the predicted values and the actual values. Lower MSE values indicate better model performance.
* **Root Mean Squared Error (RMSE):** RMSE is the square root of MSE, representing the standard deviation of the prediction errors. Similar to MSE, lower RMSE values indicate better model performance.
* **Mean Absolute Error (MAE):** MAE measures the average absolute difference between the predicted values and the actual values. It is less sensitive to outliers compared to MSE or RMSE.
* **R-squared (Coefficient of Determination):** R-squared measures the proportion of the variance in the target variable that is explained by the model. Higher R-squared values indicate better model fit.

**For Classification**

* **Accuracy:** Accuracy is the percentage of predictions that the model makes correctly. It is calculated by dividing the number of correct predictions by the total number of predictions.
* **Precision:** Precision is the percentage of positive predictions that the model makes that are actually correct. It is calculated by dividing the number of true positives by the total number of positive predictions.
* **Recall:** Recall is the percentage of all positive examples that the model correctly identifies. It is calculated by dividing the number of true positives by the total number of positive examples.
* **F1 score:** The F1 score is a weighted average of precision and recall. It is calculated by taking the harmonic mean of precision and recall.
* **Confusion matrix:** A confusion matrix is a table that shows the number of predictions for each class, along with the actual class labels. It can be used to visualize the performance of the model and identify areas where the model is struggling.

**Applications of Supervised learning**

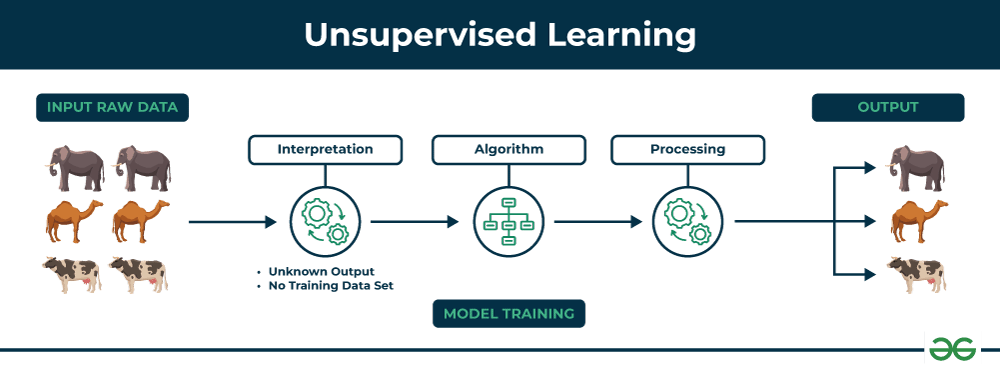
Supervised learning can be used to solve a wide variety of problems, including:

* **Spam filtering:** Supervised learning algorithms can be trained to identify and classify spam emails based on their content, helping users avoid unwanted messages.
* **Image classification:** Supervised learning can automatically classify images into different categories, such as animals, objects, or scenes, facilitating tasks like image search, content moderation, and image-based product recommendations.
* **Medical diagnosis:** Supervised learning can assist in medical diagnosis by analyzing patient data, such as medical images, test results, and patient history, to identify patterns that suggest specific diseases or conditions.
* **Fraud detection:** Supervised learning models can analyze financial transactions and identify patterns that indicate fraudulent activity, helping financial institutions prevent fraud and protect their customers.
* **Natural language processing (NLP):** Supervised learning plays a crucial role in NLP tasks, including sentiment analysis, machine translation, and text summarization, enabling machines to understand and process human language effectively.

**What is Unsupervised learning?**

Unsupervised learning is a type of machine learning that learns from unlabeled data. This means that the data does not have any pre-existing labels or categories. The goal of unsupervised learning is to discover patterns and relationships in the data without any explicit guidance.

Unlike supervised learning, no teacher is provided that means no training will be given to the machine. Therefore the machine is restricted to find the hidden structure in unlabeled data by itself.



**Types of Unsupervised Learning**

Unsupervised learning is classified into two categories of algorithms:

* **Clustering**: A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.
* **Association**: An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

**Clustering**

Clustering is a type of unsupervised learning that is used to group similar data points together. [Clustering algorithms](https://www.geeksforgeeks.org/clustering-in-machine-learning/) work by iteratively moving data points closer to their cluster centers and further away from data points in other clusters.

1. Exclusive (partitioning)
2. Agglomerative
3. Overlapping
4. Probabilistic

**Clustering Types:-**

1. Hierarchical clustering
2. K-means clustering
3. Principal Component Analysis
4. Singular Value Decomposition
5. Independent Component Analysis
6. Gaussian Mixture Models (GMMs)
7. Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

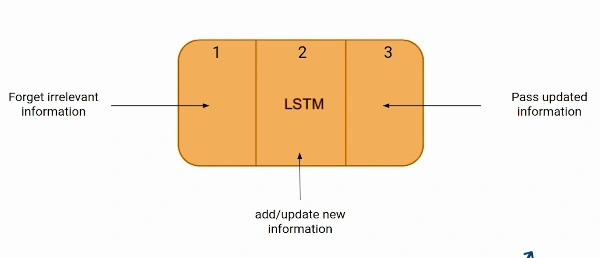
| **Parameters** | **Supervised machine learning** | **Unsupervised machine learning** |
| --- | --- | --- |
| **Input Data** | Algorithms are trained using labeled data. | Algorithms are used against data that is not labeled |
| **Computational Complexity** | Simpler method | Computationally complex |
| **Accuracy** | Highly accurate | Less accurate |
| **No. of classes** | No. of classes is known | No. of classes is not known |
| **Data Analysis** | Uses offline analysis | Uses real-time analysis of data |
| **Algorithms used** | Linear and Logistics regression,KNN Random forest, multi-class classification, decision tree, Support Vector Machine, Neural Network, etc. | K-Means clustering, Hierarchical clustering, Apriori algorithm, etc. |
| **Output** | Desired output is given. | Desired output is not given. |
| **Training data** | Use training data to infer model. | No training data is used. |
| **Complex model** | It is not possible to learn larger and more complex models than with supervised learning. | It is possible to learn larger and more complex models with unsupervised learning. |
| **Model** | We can test our model. | We can not test our model. |
| **Called as** | Supervised learning is also called classification. | Unsupervised learning is also called clustering. |
| **Example** | Example: Optical character recognition. | Example: Find a face in an image. |
| **Supervision** | supervised learning needs supervision to train the model. | Unsupervised learning does not need any supervision to train the model. |

What is LSTM?

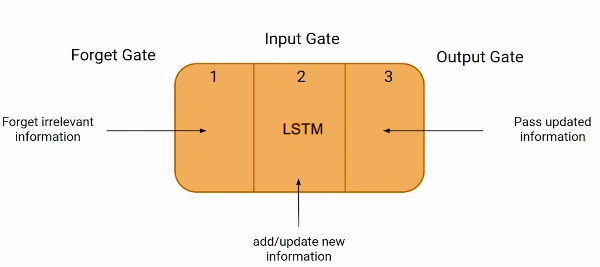
LSTM (Long Short-Term Memory) is a recurrent neural network (RNN) architecture widely used in Deep Learning. It excels at capturing long-term dependencies, making it ideal for sequence prediction tasks.

LSTM Architecture

In the introduction to long short-term memory, we learned that it resolves the vanishing gradient problem faced by RNN, so now, in this section, we will see how it resolves this problem by learning the architecture of the LSTM. At a high level, LSTM works very much like an RNN cell. Here is the internal functioning of the LSTM network. The LSTM network architecture consists of three parts, as shown in the image below, and each part performs an individual function.



These three parts of an LSTM unit are known as gates. They control the flow of information in and out of the memory cell or lstm cell. The first gate is called Forget gate, the second gate is known as the Input gate, and the last one is the Output gate. An LSTM unit that consists of these three gates and a memory cell or lstm cell can be considered as a layer of neurons in traditional feedforward neural network, with each neuron having a hidden layer and a current state.



**Linear Regression:**

Linear regression is a type of [supervised machine learning](https://www.geeksforgeeks.org/supervised-machine-learning/) algorithm that computes the linear relationship between the dependent variable and one or more independent features by fitting a linear equation to observed data.

When there is only one independent feature, it is known as [Simple Linear Regression](https://www.geeksforgeeks.org/simple-linear-regression-using-r/), and when there are more than one feature, it is known as [Multiple Linear Regression](https://www.geeksforgeeks.org/ml-multiple-linear-regression-using-python/).

**Simple Linear Regression**

This is the simplest form of linear regression, and it involves only one independent variable and one dependent variable. The equation for simple linear regression is:  
y=β0+β1X*y*=*β*0​+*β*1​*X*  
where:

* Y is the dependent variable
* X is the independent variable
* β0 is the intercept
* β1 is the slope

**Multiple Linear Regression**

This involves more than one independent variable and one dependent variable. The equation for multiple linear regression is:  
y=β0+β1X1+β2X2+………βnXn*y*=*β*0​+*β*1​*X*1+*β*2​*X*2+………*βn*​*Xn*  
where:

* Y is the dependent variable
* X1, X2, …, Xn are the independent variables
* β0 is the intercept
* β1, β2, …, βn are the slopes

**The goal of the algorithm is to find the best Fit Line equation that can predict the values based on the independent variables.**

**What is the best Fit Line?**

Our primary objective while using linear regression is to locate the best-fit line, which implies that the error between the predicted and actual values should be kept to a minimum. There will be the least error in the best-fit line.

**Hypothesis function in Linear Regression**

As we have assumed earlier that our independent feature is the experience i.e X and the respective salary Y is the dependent variable. Let’s assume there is a linear relationship between X and Y then the salary can be predicted using:

Y^=θ1+θ2X*Y*^=*θ*1​+*θ*2​*X*

OR

y^i=θ1+θ2xi*y*^​*i*​=*θ*1​+*θ*2​*xi*​

Here,

* yiϵY(i=1,2,⋯,n)     *yi*​*ϵY*(*i*=1,2,⋯,*n*)  are labels to data (Supervised learning)
* xiϵX(i=1,2,⋯,n)     *xi*​*ϵX*(*i*=1,2,⋯,*n*)  are the input independent training data (univariate – one input variable(parameter))
* yi^ϵY^(i=1,2,⋯,n)     *yi*​^​*ϵY*^(*i*=1,2,⋯,*n*)  are the predicted values.

## Cost function for Linear Regression

The [cost function](https://www.geeksforgeeks.org/what-is-cost-function/) or the[loss function](https://www.geeksforgeeks.org/ml-common-loss-functions/) is nothing but the error or difference between the predicted value Y^     *Y*^  and the true value Y.

In Linear Regression, the **Mean Squared Error (MSE)** cost function is employed, which calculates the average of the squared errors between the predicted values y^i*y*^​*i*​ and the actual values yi*yi*​. The purpose is to determine the optimal values for the intercept θ1*θ*1​ and the coefficient of the input feature θ2*θ*2​ providing the best-fit line for the given data points. The linear equation expressing this relationship is y^i=θ1+θ2xi*y*^​*i*​=*θ*1​+*θ*2​*xi*​.

MSE function can be calculated as:

Cost function(J)=1n∑ni(yi^−yi)2Cost function(*J*)=*n*1​∑*ni*​(*yi*​^​−*yi*​)2

**Gradient Descent for Linear Regression**

A linear regression model can be trained using the optimization algorithm [gradient descent](https://www.geeksforgeeks.org/gradient-descent-algorithm-and-its-variants/)by iteratively modifying the model’s parameters to reduce the[mean squared error (MSE)](https://www.geeksforgeeks.org/python-mean-squared-error/) of the model on a training dataset.

[IPL 2024 RCB vs DC Analysis using Python | Aman Kharwal](https://thecleverprogrammer.com/2024/05/20/ipl-2024-rcb-vs-dc-analysis-using-python/)