

Introduction to Data Science

Dr. Irfan Yousuf

Department of Computer Science (New Campus)

UET, Lahore

(Lecture # 15; October 24, 2022)

Outline

- Basics of Machine Learning

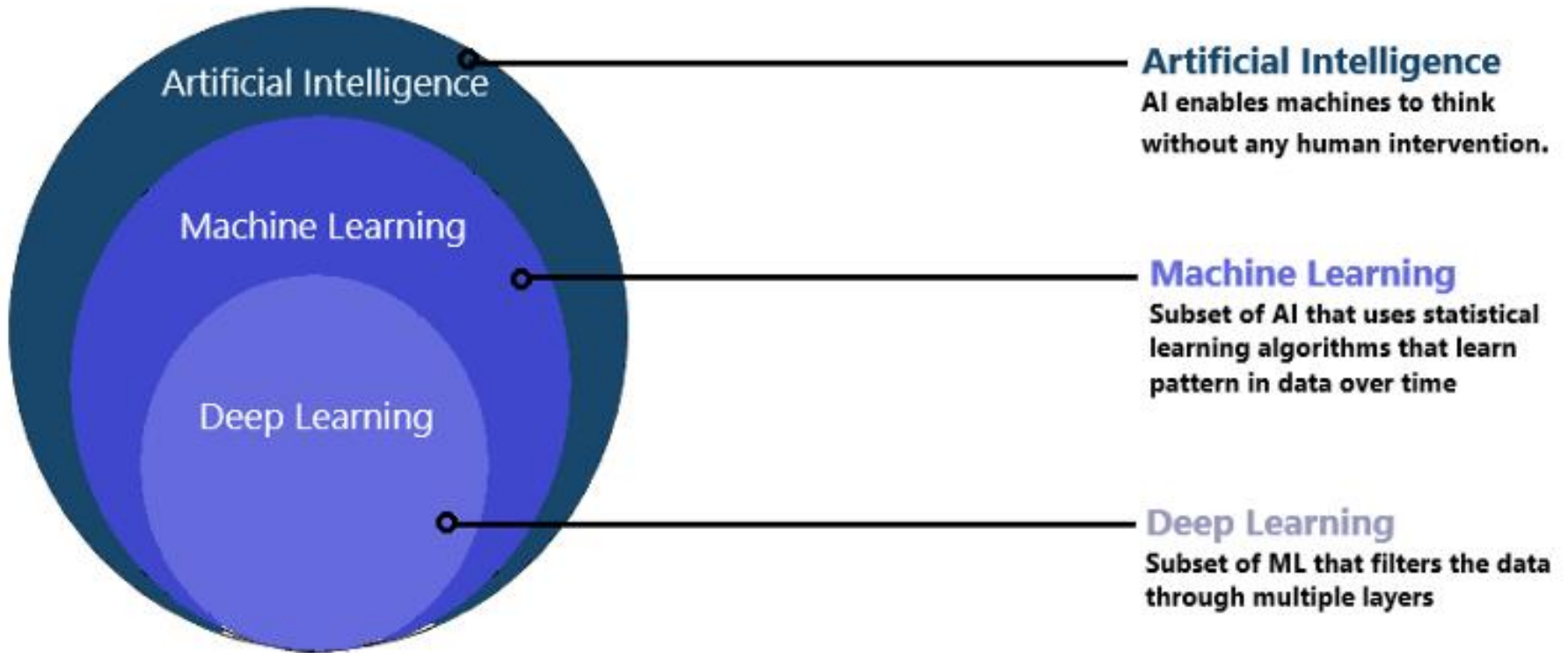
Machine Learning

- Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.
- Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves.

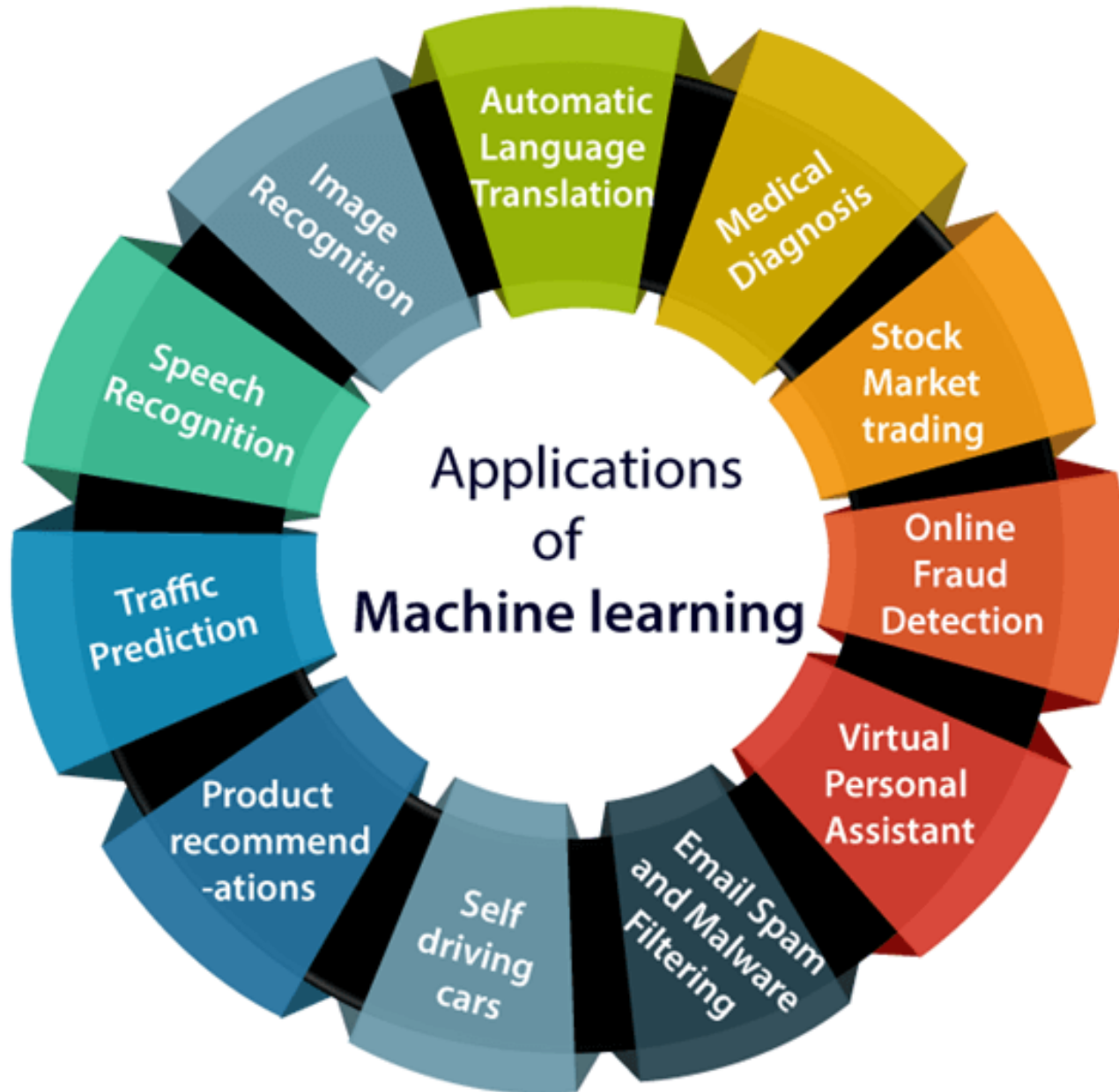
Machine Learning

- 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 (ML) is the study of computer algorithms that improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence.
- In Machine Learning, we train a machine **how** to learn.

Machine Learning



Applications of Machine Learning



Types of Machine Learning Algorithms

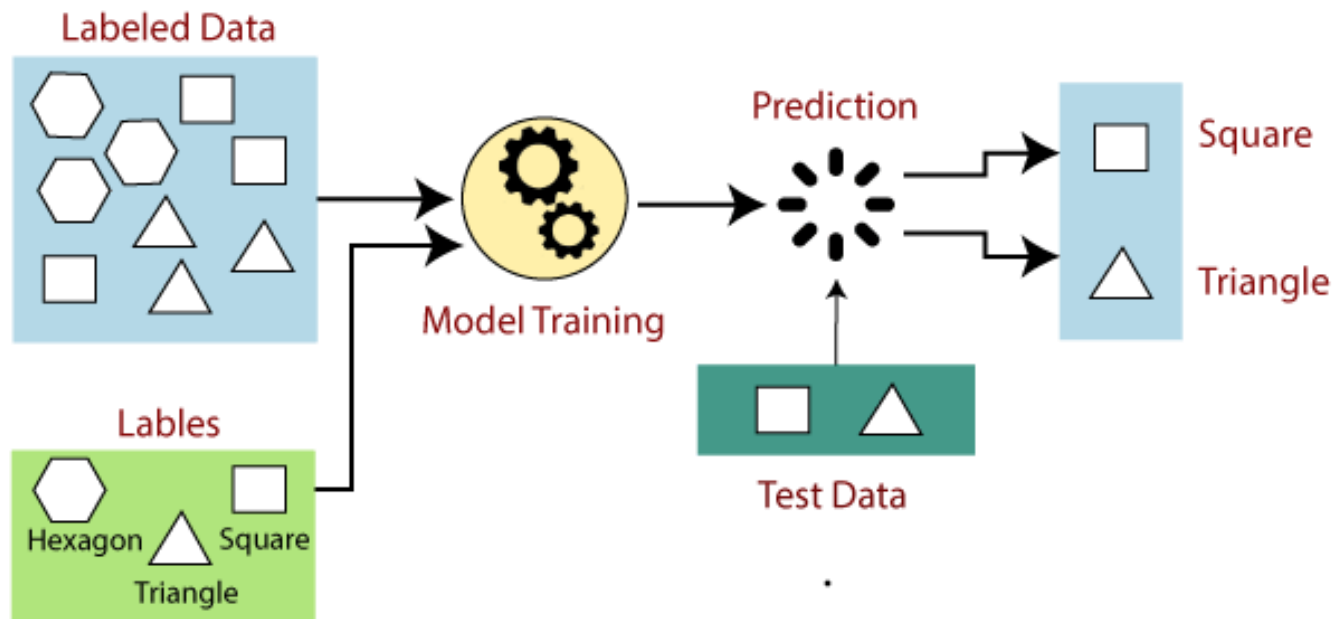
- Supervised Machine Learning
- Un-supervised Machine Learning
- Semi-supervised Machine Learning
- Reinforcement Learning

Supervised Learning

- Algorithms are trained using labeled examples, such as an input where the desired output is known.
- Can apply what has been learned in the past to new data using labeled examples to predict future events.
- Supply algorithms with labeled training data and define the variables they want the algorithm to assess for correlations.
- **In machine learning, a target is called a label.**
- **In statistics, a target is called a dependent variable.**

Supervised Learning

- Through methods like classification, regression, prediction and gradient boosting, supervised learning uses patterns to predict the values of the label on additional unlabeled data.
- Supervised learning is commonly used in applications where historical data predicts likely future events

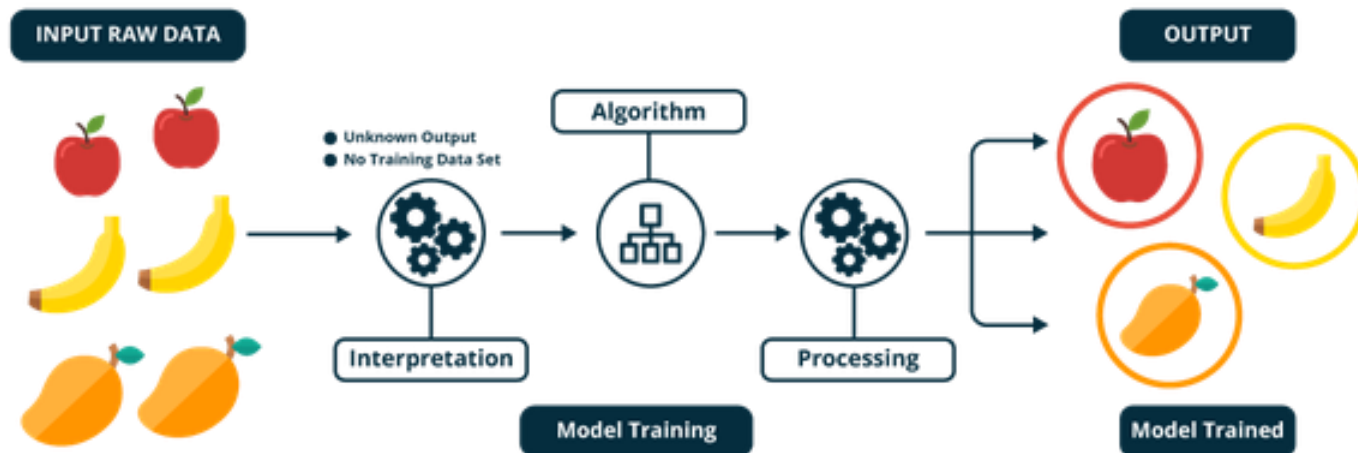


Un-supervised Learning

- Unsupervised learning is used against data that has no historical labels. The system is not told the "right answer." The algorithm must figure out what is being shown. The goal is to explore the data and find some structure within.
- Unsupervised learning studies how systems can infer a function to **describe a hidden structure from unlabeled data**. The system doesn't figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.
- The algorithm scans through data sets looking for any meaningful connection.

Un-supervised Learning

- Unsupervised learning works well on transactional data. For example, it can identify segments of customers with similar attributes who can then be treated similarly in marketing campaigns.
- **A variable in statistics is called a feature in machine learning.**

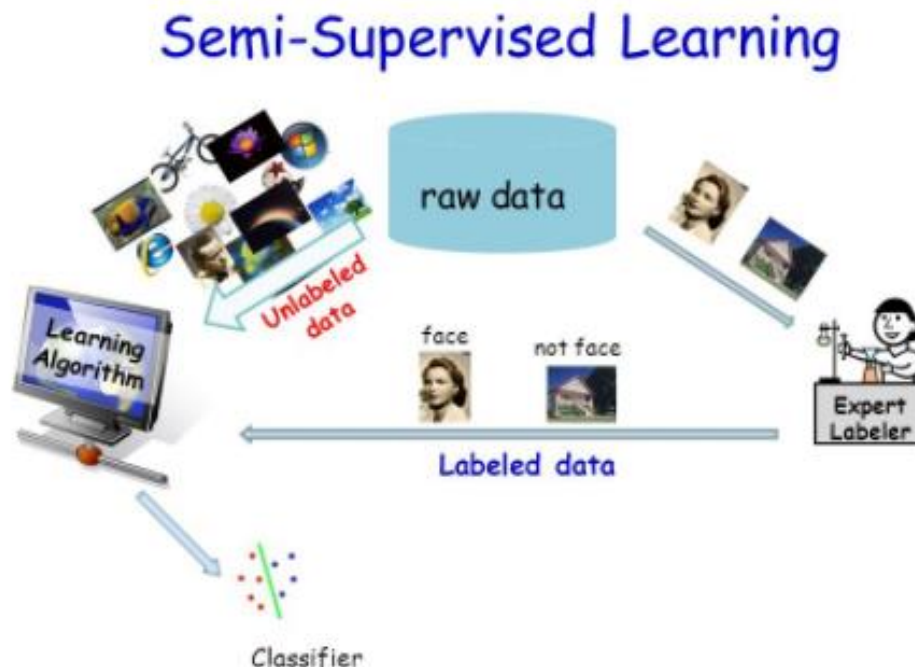


Semi-supervised Learning

- It is used for the same applications as supervised learning. But it uses both labeled and unlabeled data for training – typically a small amount of labeled data with a large amount of unlabeled data (because unlabeled data is less expensive and takes less effort to acquire).
- This type of learning can be used with methods such as classification, regression and prediction. Semi-supervised learning is useful when the cost associated with labeling is too high to allow for a fully labeled training process.

Semi-supervised Learning

- Usually, semi-supervised learning is chosen when the acquired labeled data requires skilled and relevant resources in order to train it / learn from it.
- Early examples of this include identifying a person's face on a web cam.

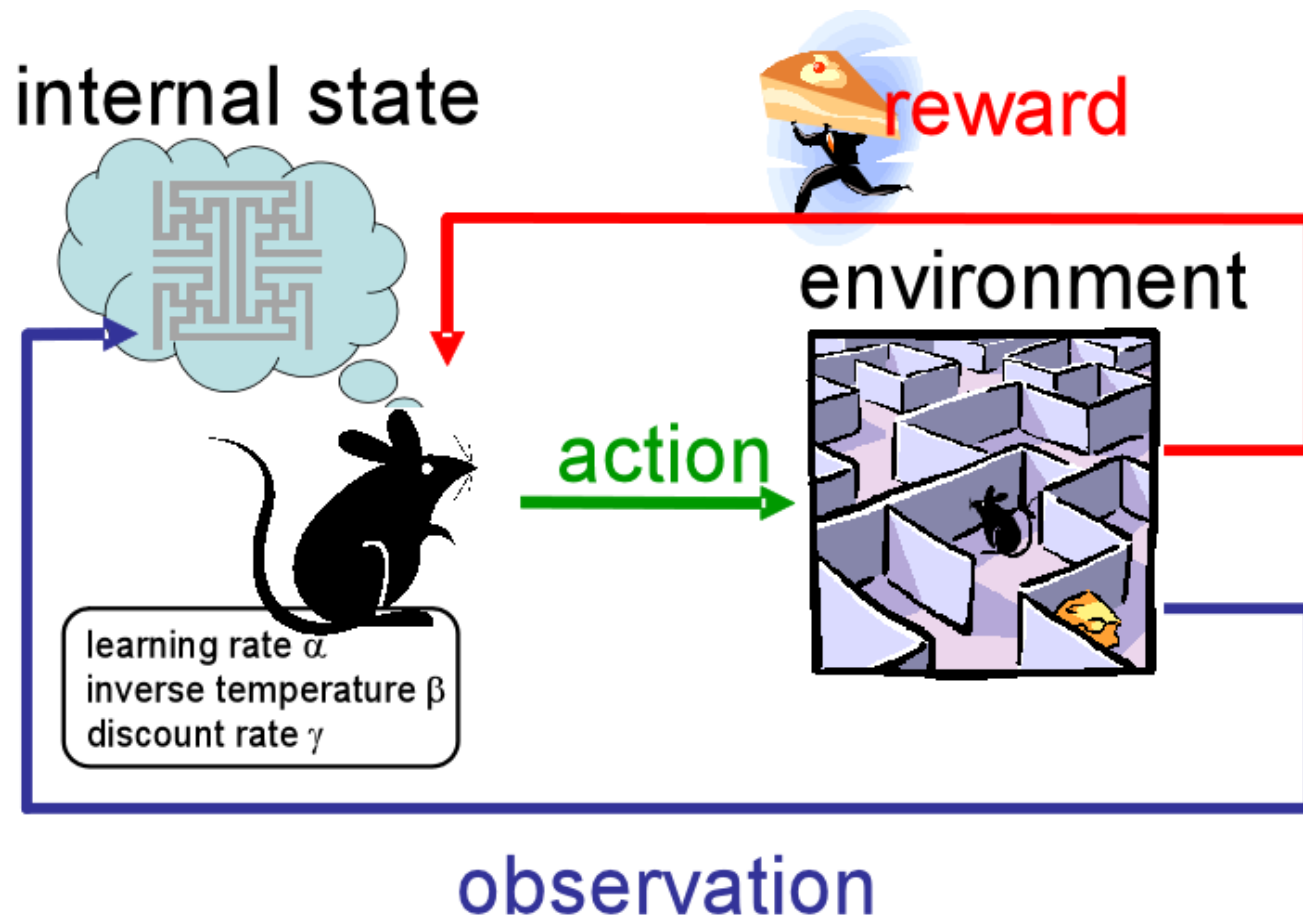


Reinforcement Learning

- With reinforcement learning, the algorithm discovers through trial and error which actions yield the greatest rewards. This type of learning has three primary components: the agent (the learner or decision maker), the environment (everything the agent interacts with) and actions (what the agent can do).
- Data scientists program an algorithm to complete a task and give it positive or negative cues as it works out how to complete a task. But for the most part, the algorithm decides on its own what steps to take along the way.

Reinforcement Learning

- The objective is for the agent to choose actions that maximize the expected reward over a given amount of time.



Machine Learning Algorithms

Machine Learning

Supervised learning: Train a model with known input and output data to predict future outputs to new data.

Classification

Support vector machine (SVM)

K-nearest-neighbors

Discriminant analysis

Neural Networks

Naive Bayes

Regression

Linear Regression

Assembly Methods

Decision trees

Neural Networks

Unsupervised Learning: Segment a collection of elements with the same attributes (clustering).

Clustering

K-means, k-medoids fuzzy C-means

Hidden Markov models

Neural Networks

Gaussian mixture

Supervised Machine Learning Algorithms

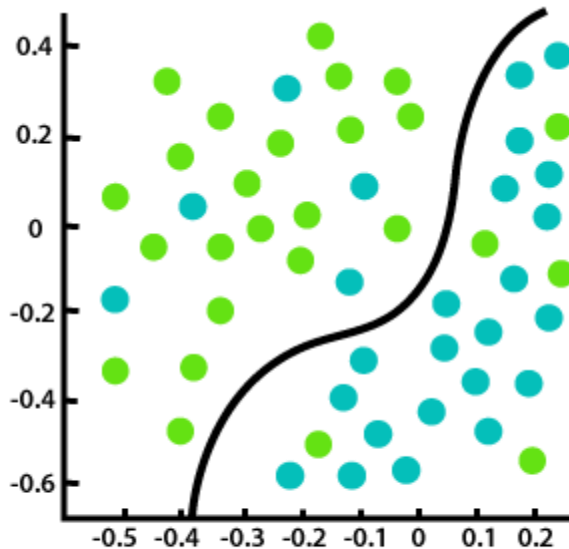
- Supervised learning is where you have input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output.

$$Y = f(X)$$

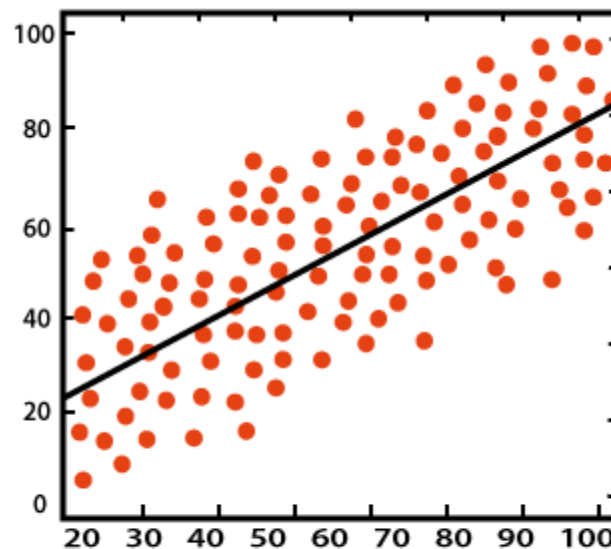
- The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.
- The process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process. We know the correct answers, the algorithm iteratively makes predictions on the training data and is corrected by the teacher. Learning stops when the algorithm achieves an acceptable level of performance.

Supervised Machine Learning Algorithms

- **Classification:** A classification problem is when the output variable is a category, such as “red” or “blue” or “disease” and “no disease”.
- **Regression:** A regression problem is when the output variable is a real value, such as “age” or “weight”.



Classification



Regression

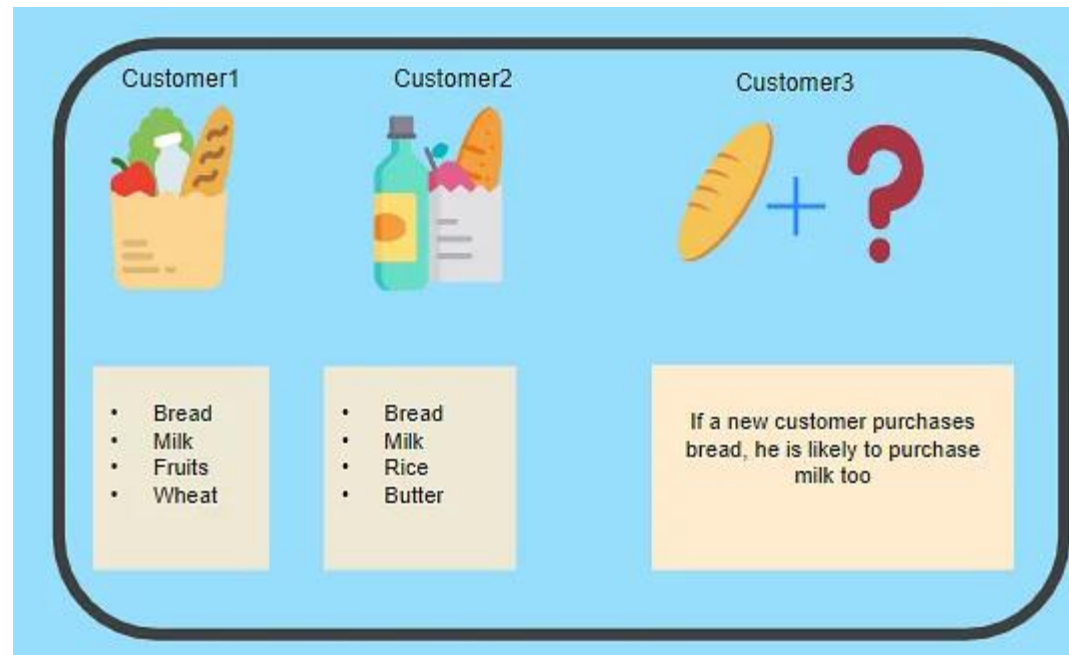
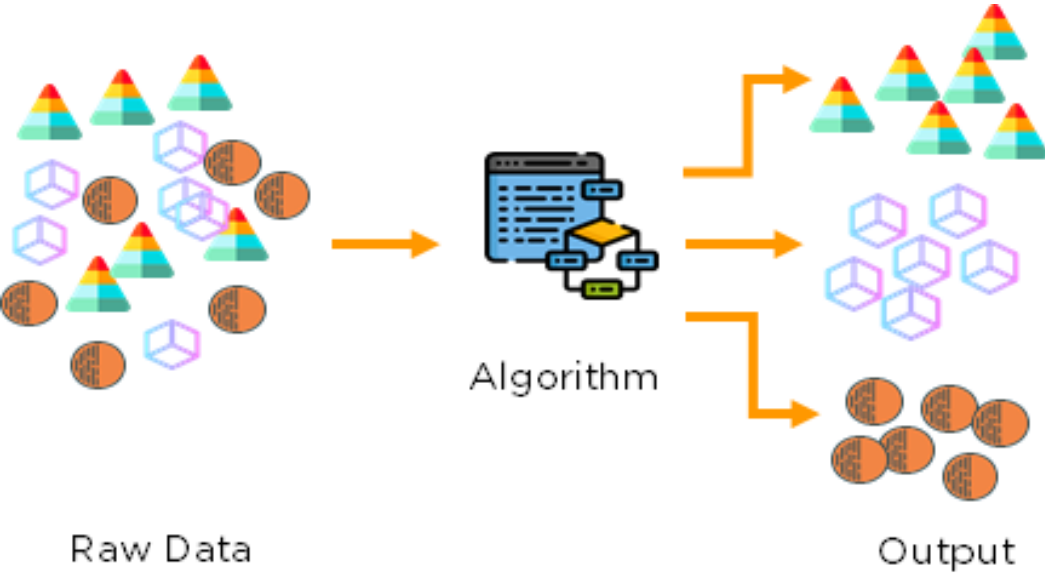
Unsupervised Machine Learning Algorithms

- Unsupervised learning refers to the use of artificial intelligence (AI) algorithms to identify patterns in data sets containing data points that are not labeled.
- The algorithms are thus allowed to classify, label and/or group the data points contained within the data sets without having any external guidance in performing that task.
- In unsupervised learning, an AI system will group unsorted information according to similarities and differences even though there are no categories provided.

Unsupervised Machine Learning Algorithms

- **Clustering:** Clustering is a method of grouping the objects into clusters such that objects with most similarities remains into a group and has less or no similarities with the objects of another group.
- **Association:** An association rule is an unsupervised learning method which is used for finding the relationships between variables in the large database. It determines the set of items that occurs together in the dataset. Association rule makes marketing strategy more effective. Such as people who buy X item (suppose a bread) are also tend to purchase Y (Butter/Jam) item.

Unsupervised Machine Learning Algorithms



Summary

- Basics of Machine Learning