Unsupervised Machine Learning: What is, Algorithms, Example

What is Unsupervised Learning?

Unsupervised learning is a machine learning technique, where you do not need to supervise the model. Instead, you need to allow the model to work on its own to discover information. It mainly deals with the unlabelled data.

Unsupervised learning algorithms allows you to perform more complex processing tasks compared to supervised learning. Although, unsupervised learning can be more unpredictable compared with other natural learning methods.

In this tutorial, you will learn:

Example of Unsupervised Machine Learning

Let's, take the case of a baby and her family dog.

(/images/1/030819_1030_Unsupervise1.png)

She knows and identifies this dog. Few weeks later a family friend brings along a dog and tries to play with the baby.

(/images/1/030819_1030_Unsupervise2.png)

Baby has not seen this dog earlier. But it recognizes many features (2 ears, eyes, walking on 4 legs) are like her pet dog. She identifies the new animal as a dog. This is unsupervised learning, where you are not taught but you learn from the data (in this case data about a dog.) Had this been supervised learning, the family friend would have told the baby that it's a dog.

Why Unsupervised Learning?

Here, are prime reasons for using Unsupervised Learning:

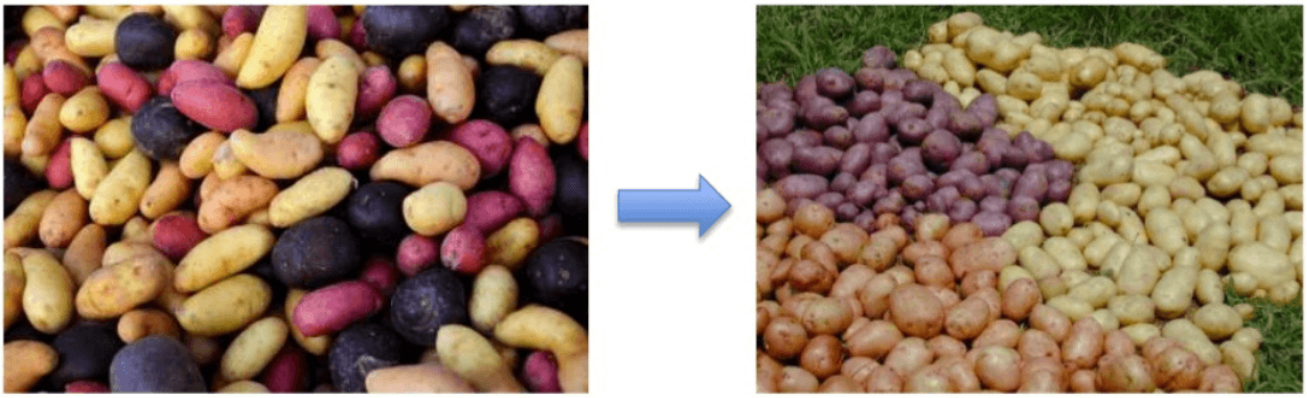
- Unsupervised machine learning finds all kind of unknown patterns in data.
- Unsupervised methods help you to find features which can be useful for categorization.

- It is taken place in real time, so all the input data to be analyzed and labeled in the presence of learners.
- It is easier to get unlabeled data from a computer than labeled data, which needs manual intervention.

Types of Unsupervised Learning

Unsupervised learning problems further grouped into clustering and association problems.

Clustering



(/images/1/030819_1030_Unsupervise3.png)

Clustering is an important concept when it comes to unsupervised learning. It mainly deals with finding a structure or pattern in a collection of uncategorized data. Clustering algorithms will process your data and find natural clusters(groups) if they exist in the data. You can also modify how many clusters your algorithms should identify. It allows you to adjust the granularity of these groups.

There are different types of clustering you can utilize:

Exclusive (partitioning)

In this clustering method, Data are grouped in such a way that one data can belong to one cluster only.

Example: K-means

Agglomerative

In this clustering technique, every data is a cluster. The iterative unions between the two nearest clusters reduce the number of clusters.

Example: Hierarchical clustering

Overlapping

In this technique, fuzzy sets is used to cluster data. Each point may belong to two or more clusters with separate degrees of membership.

Here, data will be associated with an appropriate membership value. Example: Fuzzy C-Means

Probabilistic

This technique uses probability distribution to create the clusters

Example: Following keywords

- "man's shoe."
- "women's shoe."
- "women's glove."
- "man's glove."

can be clustered into two categories "shoe" and "glove" or "man" and "women."

Clustering Types

- Hierarchical clustering
- K-means clustering
- K-NN (k nearest neighbors)
- Principal Component Analysis
- Singular Value Decomposition
- Independent Component Analysis

Hierarchical Clustering:

Hierarchical clustering is an algorithm which builds a hierarchy of clusters. It begins with all the data which is assigned to a cluster of their own. Here, two close cluster are going to be in the same cluster. This algorithm ends when there is only one cluster left.

K-means Clustering

K means it is an iterative clustering algorithm which helps you to find the highest value for every iteration. Initially, the desired number of clusters are selected. In this clustering method, you need to cluster the data points into k groups. A larger k means smaller groups

with more granularity in the same way. A lower k means larger groups with less granularity.

The output of the algorithm is a group of "labels." It assigns data point to one of the k groups. In k-means clustering, each group is defined by creating a centroid for each group. The centroids are like the heart of the cluster, which captures the points closest to them and adds them to the cluster.

K-mean clustering further defines two subgroups:

- Agglomerative clustering
- Dendrogram

Agglomerative clustering:

This type of K-means clustering starts with a fixed number of clusters. It allocates all data into the exact number of clusters. This clustering method does not require the number of clusters K as an input. Agglomeration process starts by forming each data as a single cluster.

This method uses some distance measure, reduces the number of clusters (one in each iteration) by merging process. Lastly, we have one big cluster that contains all the objects.

Dendrogram:

In the Dendrogram clustering method, each level will represent a possible cluster. The height of dendrogram shows the level of similarity between two join clusters. The closer to the bottom of the process they are more similar cluster which is finding of the group from dendrogram which is not natural and mostly subjective.

K- Nearest neighbors

K- nearest neighbour is the simplest of all machine learning classifiers. It differs from other machine learning techniques, in that it doesn't produce a model. It is a simple algorithm which stores all available cases and classifies new instances based on a similarity measure.

It works very well when there is a distance between examples. The learning speed is slow when the training set is large, and the distance calculation is nontrivial.

Principal Components Analysis:

In case you want a higher-dimensional space. You need to select a basis for that space and only the 200 most important scores of that basis. This base is known as a principal component. The subset you select constitute is a new space which is small in size compared to original space. It maintains as much of the complexity of data as possible.

Association

Association rules allow you to establish associations amongst data objects inside large databases. This unsupervised technique is about discovering interesting relationships between variables in large databases. For example, people that buy a new home most likely to buy new furniture.

Other Examples:

- A subgroup of cancer patients grouped by their gene expression measurements
- Groups of shopper based on their browsing and purchasing histories
- Movie group by the rating given by movies viewers

Supervised vs. Unsupervised Machine Learning

Parameters	Supervised machine learning technique	Unsupervised machine learning technique
Input Data	Algorithms are trained using labeled data.	Algorithms are used against data which is not labelled
Computational Complexity	Supervised learning is a simpler method.	Unsupervised learning is computationally complex
Accuracy	Highly accurate and trustworthy method.	Less accurate and trustworthy method.

Applications of unsupervised machine learning

Some applications of unsupervised machine learning techniques are:

- Clustering automatically split the dataset into groups base on their similarities
- Anomaly detection can discover unusual data points in your dataset. It is useful for finding fraudulent transactions
- Association mining identifies sets of items which often occur together in your dataset
- Latent variable models are widely used for data preprocessing. Like reducing the number of features in a dataset or decomposing the dataset into multiple components

Disadvantages of Unsupervised Learning

- You cannot get precise information regarding data sorting, and the output as data used in unsupervised learning is labeled and not known

- Less accuracy of the results is because the input data is not known and not labeled by people in advance. This means that the machine requires to do this itself.
- The spectral classes do not always correspond to informational classes.
- The user needs to spend time interpreting and label the classes which follow that classification.
- Spectral properties of classes can also change over time so you can't have the same class information while moving from one image to another.

Summary

- Unsupervised learning is a machine learning technique, where you do not need to supervise the model.
- Unsupervised machine learning helps you to finds all kind of unknown patterns in data.
- Clustering and Association are two types of Unsupervised learning.
- Four types of clustering methods are 1) Exclusive 2) Agglomerative 3) Overlapping 4) Probabilistic.
- Important clustering types are: 1)Hierarchical clustering 2) K-means clustering 3) K-NN 4) Principal Component Analysis 5) Singular Value Decomposition 6) Independent Component Analysis.
- Association rules allow you to establish associations amongst data objects inside large databases.
- In Supervised learning, Algorithms are trained using labelled data while in Unsupervised learning Algorithms are used against data which is not labelled.
- Anomaly detection can discover important data points in your dataset which is useful for finding fraudulent transactions.
- The biggest drawback of Unsupervised learning is that you cannot get precise information regarding data sorting.

Prev (/supervised-machine-learning.html)

Report a Bug

Next (/supervised-vs-unsupervised-learning.html)

YOU MIGHT LIKE:

DATA WAREHOUSING

(/data-mining-tutorial.html) (/data-

DATA WAREHOUSING

(/datastage-tutorial.html) (/datastage-tutorial.html)

DATA WAREHOUSING

(/business-intelligence-definition-example.html)

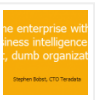
mining-tutorial.html)

Data Mining Tutorial: Process, Techniques, Tools & Applications

(/data-mining-tutorial.html)

DataStage Tutorial: Beginner's Training

(/datastage-tutorial.html)



(/business-intelligence-definition-example.html)

What is Business Intelligence? Definition & Example

(/business-intelligence-definition-example.html)

DATA WAREHOUSING

(/database-vs-data-warehouse.html)

(/database-vs-data-warehouse.html)

Database vs Data Warehouse: Key Differences

(/database-vs-data-warehouse.html)

DATA WAREHOUSING

(/etl-testing-tools.html)

(/etl-testing-tools.html)

5 Best ETL Automation Testing Tools in 2019

(/etl-testing-tools.html)

TABLEAU

(/tableau-sort-data.html)

(/tableau-sort-data.html)

Tableau Create Group, Hierarchy, Sets & Sort Data

(/tableau-sort-data.html)

AI Tutorial

1) What is AI? (/artificial-intelligence-tutorial.html)

2) What is Machine Learning? (/machine-learning-tutorial.html)

3) Deep learning Tutorial (/deep-learning-tutorial.html)

4) Machine Learning vs Deep Learning (/machine-learning-vs-deep-learning.html)

5) Supervised Machine Learning (/supervised-machine-learning.html)

6) Unsupervised Machine Learning (/unsupervised-machine-learning.html)

7) Supervised vs Unsupervised Learning (/supervised-vs-unsupervised-learning.html)

8) Back Propagation Neural Network (/backpropogation-neural-network.html)

9) Deep Learning Libraries (/deep-learning-libraries.html)

10) Fuzzy Logic Tutorial (/what-is-fuzzy-logic.html)

11) AI & Machine Learning Books (/ai-machine-learning-books.html)

(https://www.facebook.com/guru99com/) (https://twitter.com/guru99com)

(https://www.youtube.com/channel/UC19i1XD6k88KqHlET8atqFQ)

(https://forms.aweber.com/form/46/724807646.htm)

About

About Us (/about-us.html)
Advertise with Us (/advertise-us.html)
Write For Us (/become-an-instructor.html)
Contact Us (/contact-us.html)

Career Suggestion

SAP Career Suggestion Tool (/best-sap-module.html)
Software Testing as a Career (/software-testing-career-complete-guide.html)
Certificates (/certificate-it-professional.html)

Interesting

Books to Read! (/books.html)
Blog (/blog/)
Quiz (/tests.html)
eBook (/ebook-pdf.html)

Execute online

Execute Java Online (/try-java-editor.html)
Execute Javascript (/execute-javascript-online.html)
Execute HTML (/execute-html-online.html)
Execute Python (/execute-python-online.html)

Privacy Policy (/privacy-policy.html)