

Machine Learning

Dr Changjiang He, Dr Kuo-Ming Chao
Computer Science| School of Art
University of Roehampton

Lesson 7.1

Unsupervised Learning

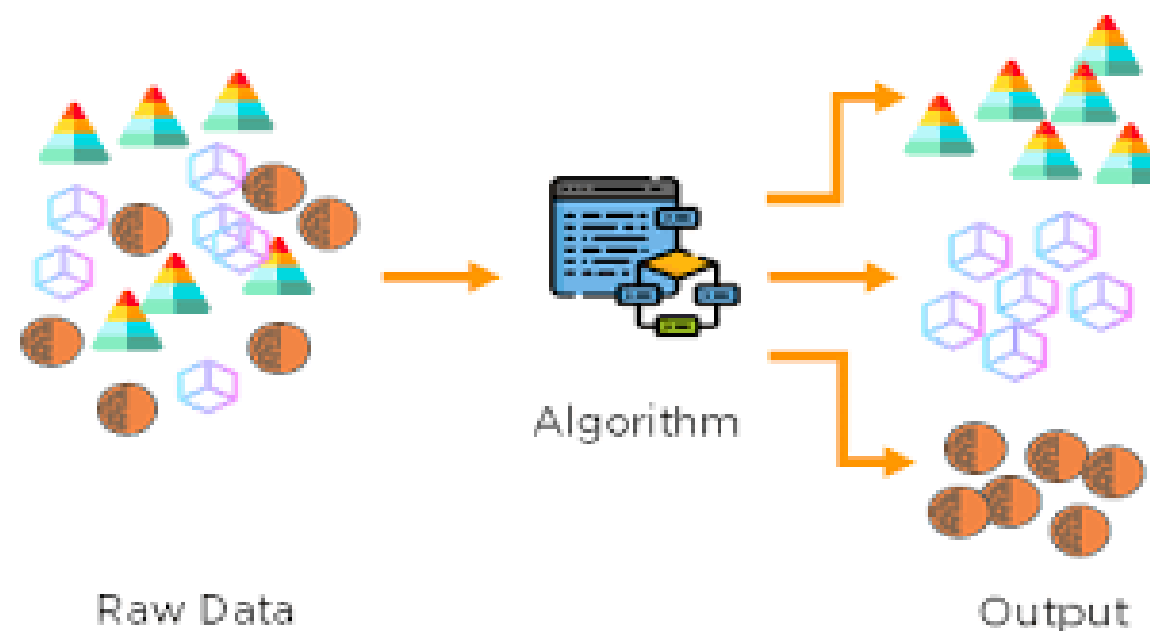
- What is unsupervised learning?
- Types of unsupervised learning
- Supervised learning vs unsupervised learning

- Unlike supervised machine learning, unsupervised machine learning methods cannot be directly applied to a regression or a classification problem.
- Because we have no idea what the values for the output data might be, making it impossible for us to train the algorithm the way we normally would.

In simple word, we only have input but do not have output or label

What is Unsupervised Learning?

- Unsupervised machine learning algorithms infer patterns from a dataset without reference to known, or labelled, outcomes.
- Unsupervised learning can instead be used to discover the underlying structure of the data.



What is Unsupervised Learning?

- For instance, suppose it is given an image having both dogs and cats which it has never seen.



What is Unsupervised Learning?

- Thus the machine has no idea about the features of dogs and cats so we can't categorize it as 'dogs and cats '.
- But it can categorize them according to their similarities, patterns, and differences, i.e., we can easily categorize the above picture into two parts.
- The first may contain all pics having **dogs** in them and the second part may contain all pics having **cats** in them.

It allows the model to work on its own to discover patterns and information that was previously undetected.

It mainly deals with unlabelled data. 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 behaviour.
- **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.

Supervised Learning vs Unsupervised Learning

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, Random forest, Support Vector Machine, Neural Network, etc.	K-Means clustering, Hierarchical clustering, Apriori algorithm, etc.

Some clustering methods include,

- Hierarchical clustering
- K-means clustering
- Principal Component Analysis
- Singular Value Decomposition
- Independent Component Analysis