Unsupervised machine learning

1/3



Unsupervised machine learning is a type of machine learning where the algorithm learns patterns and structures in the data without being provided with labeled outcomes or target variables. In other words, it works with data that does not have predefined labels or classifications, and the goal is to identify hidden patterns, structures, or groupings within the dataset.

Key Characteristics of Unsupervised Machine Learning: 1. **No Labeled Data**: Unlike supervised learning, where the data comes with labels (e.g., a dataset with both input features and known outputs), unsupervised learning algorithms work with input data alone, without any specific target or output values.

2. **Exploration of Data**: The primary objective is to explore the data, discover the underlying structure, or group similar data points together. It's often used for tasks like clustering, dimensionality reduction, anomaly detection, and feature extraction.

3. **Output**: The output is usually a grouping of data points (clusters), reduced dimensions, or hidden relationships that help in understanding the structure of the data.

Common Techniques in Unsupervised Machine Learning: - **Clustering**: Grouping data points that are similar to each other into clusters. Popular algorithms include K-means, hierarchical clustering, and DBSCAN. - **Dimensionality Reduction**: Reducing the number of features in the data while preserving its structure, making it easier to analyze and visualize. Techniques include Principal Component Analysis (PCA) and t-SNE. - **Association Rule Learning**: Identifying interesting relationships or associations between variables in large datasets. A classic example is market basket analysis. - **Anomaly Detection**: Identifying rare or unusual data points that do not fit the general patterns, often used for fraud detection or network security.

Applications: - **Customer Segmentation**: In marketing, unsupervised learning can identify different customer groups based on purchasing behavior without predefined categories. - **Image Compression**: By reducing the number of features in an image, unsupervised techniques can be used for compression. - **Anomaly Detection**: Detecting fraudulent transactions, system malfunctions, or any outlier behavior in data.

In summary, unsupervised machine learning is ideal for analyzing datasets where the structure is not immediately known and where the objective is to identify meaningful patterns, groupings, or features from the data itself.

In this algorithm, we do not have any target or outcome variable to predict / estimate. It is used for clustering population in different groups, which is widely used for segmenting customers in different groups for specific intervention. Examples of Unsupervised Learning: Apriori algorithm, K-means.

For unsupervised learning, no prelabeling is required. These algorithms cluster data points based on similarities in features and can be powerful tools for detecting previously unknown patters in multidimensional data ¹⁾.

The progress in this field of applied machine learning (ML) is continuously driven by the growing amount of available data and the increasing computational power^{2) 3)}.

To assess the sensitivity and specificity of arteriovenous malformation (AVM) nidus component identification and quantification using unsupervised machine learning algorithm, and to evaluate the association between intervening nidal brain parenchyma and radiation-induced changes (RICs) after stereotactic radiosurgery (SRS).

Fully automated segmentation via unsupervised classification with fuzzy c-means clustering was used to analyze AVM nidus on T2-weighted magnetic resonance imaging. The proportions of vasculature, brain parenchyma, and cerebrospinal fluid (CSF) were quantified. This was compared to manual segmentation. Association between brain parenchyma component and RIC development was assessed.

The proposed algorithm was applied to 39 unruptured AVMs. This included 17 female and 22 male patients with a median age of 27 years. The median percentages of the constituents were as follows: vasculature (31.3%), brain parenchyma (48.4%), and CSF (16.8%). RICs were identified in 17 (43.6%) of 39 patients. Compared to manual segmentation, the automated algorithm was able to achieve a Dice similarity index of 79.5% (sensitivity=73.5% and specificity=85.5%). RICs were associated with higher proportions of intervening nidal brain parenchyma (52.0% vs. 45.3%, p=0.015). Obliteration was not associated with a higher proportions of nidal vasculature (36.0% vs. 31.2%, p=0.152).

The automated segmentation algorithm was able to achieve classification of AVM nidus components with relative accuracy. Higher proportions of intervening nidal brain parenchyma were associated with RICs $^{4)}$.

1)

Deo RC. Machine learning in medicine. Circulation . 2015;132(20):1920-1930.

Rodrigues JFJr, Paulovich FV, de Oliveira MC, de Oliveira ONJr. On the convergence of nanotechnology and Big Data analysis for computer-aided diagnosis. Nanomedicine (Lond) . 2016;11(8):959-982.

Coveney PV, Dougherty ER, Highfield RR. Big data need big theory too. Philos Trans A Math Phys Eng Sci . 2016;374(2080).

Lee CC, Yang HC, Lin CJ, Chen CJ, Wu HM, Shiau CY, Guo WY, Hung-Chi Pan D, Liu KD, Chung WY, Peng SJ. Intervening nidal brain parenchyma and risk of radiation-induced changes after radiosurgery for brain arteriovenous malformation: a study using unsupervised machine learning algorithm. World Neurosurg. 2019 Jan 21. pii: S1878-8750(19)30103-2. doi: 10.1016/j.wneu.2018.12.220. [Epub ahead of print] PubMed PMID: 30677586.

Permanent link: https://neurosurgerywiki.com/wiki/doku.php?id=unsupervised_machine_learning



