



DBSCAN Algorithm in Machine Learning

Introduction to DBSCAN Algorithm in Machine Learning.



AMAN KHARWAL / ⌚ OCTOBER 31, 2020 / 📁 MACHINE LEARNING

The DBSCAN algorithm is a very useful clustering algorithm in Machine Learning. DBSCAN stands for Density-Based Spatial Clustering of Applications with Noise. In this article, I will take you through what is the DBSCAN algorithm in Machine Learning and how it works.

M1 Pro

techscroll.in

Introduction to DBSCAN Algorithm in Machine Learning

The main advantages of DBSCAN are that it does not require the user to define the number of clusters a priori, it can capture clusters of complex shapes and it can identify points that are not part of any cluster. DBSCAN is a bit slower than agglomerate clustering and k-means but still accommodates relatively large datasets.

[Also, Read – Machine Learning Full Course For free.](#)

DBSCAN works by identifying points that are in “crowded” regions of feature space, where many data points are close to

each other. These regions are called dense regions in feature space. The idea behind DBSCAN is that clusters form dense regions of data, separated by relatively empty regions.

M1 Pro

techscroll.in

Points in a dense region are called core samples (or core points), and they are defined as follows. There are two parameters in DBSCAN: `min_samples` and `eps`. If there are many `min_samples` or more data points within a distance of `eps` to a given data point, that data point is classified as a base sample.

Base samples that are closer to each other than the `eps` distance are placed in the same cluster by DBSCAN.

How Does The DBSCAN Algorithm Work?

The DBSCAN algorithm works by choosing an arbitrary point to start. It then finds all the points with a distance `eps` or less from that point. If there are less than `min_samples` points within `eps` distance of the starting point, that point is labeled as noise, which means it does not belong to any cluster.

M1 Pro

techscroll.in

If there are more `min_samples` points within a distance of `eps`, the point is labeled as a base sample and assigned a new

cluster label. Then all the neighbours (in ϵ) of the point are visited. If they have not yet received a cluster, they receive the newly created cluster label.

If these are basic samples, their neighbours are visited in turn, and so on. The cluster grows until there are no more base samples within an ϵ distance of the cluster. Then another point which has not yet been visited is selected and the same procedure is repeated.

Ultimately, there are three types of points: centre points, points at an ϵ distance from centre points (called endpoints), and noise. When the DBSCAN algorithm is run multiple times on a particular dataset, the grouping of the centre points is always the same and the same points will always be labeled as noise.

M1 Pro

techscroll.in

However, an endpoint can be close to the base samples of several clusters. Therefore, membership of the endpoint cluster depends on the order in which the points are visited. Usually, there are only a few endpoints, and this slight dependence on the order of the points is not important.

When using DBSCAN, you should be careful about handling returned cluster assignments. You can learn the hands-on implementation of the DBSCAN algorithm from [here](#), which is a machine learning contact tracing task using the DBSCAN algorithm.

Hope you liked this article on DBSCAN Algorithm in Machine Learning. Please feel free to ask your valuable questions in the comments section below.



Aman Kharwal

Coder with the ♥ of a Writer || Data Scientist | Solopreneur |
Founder

ARTICLES: 1202



PREVIOUS

NEXT

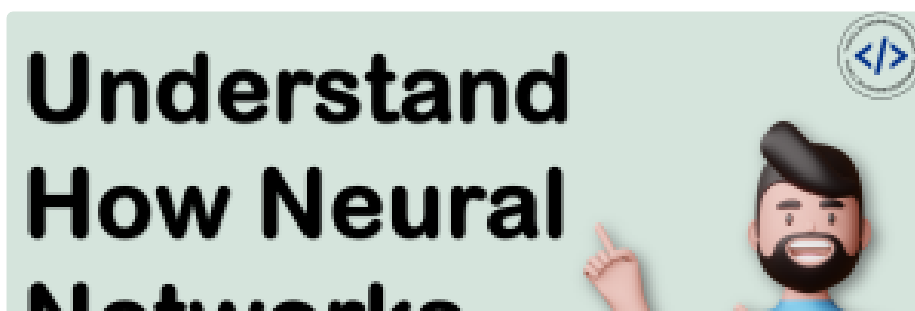


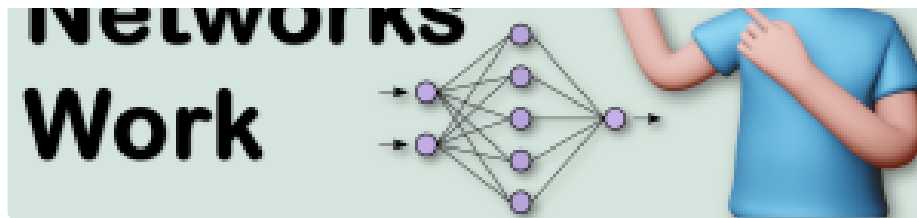
Recommended For You



Best Approaches for Time Series Analysis

January 8, 2022





How Neural Network Works

January 6, 2022



Visualize a Machine Learning Algorithm using Python

December 29, 2021



Activation Functions in Neural Networks

December 23, 2021

Leave a Reply

Enter your comment here...

 FACEBOOK  INSTAGRAM  MEDIUM  LINKEDIN

Copyright © Thecleverprogrammer.com 2022