A BRIEF INTRODUCTION TO UNSUPERVISED LEARNING

SWIPE →

01

What is unsupervised learning?

The goal of these algorithms is to analyze data and find important features

Unsupervised learning will often find subgroups or hidden patterns within the dataset that a human observer may not pick up on

02

Clustering—the most common application

Clustering aims to discover “clusters”, or subgroups within unlabeled data

Clusters contain data points that are as similar as possible to each other, and as dissimilar to data points in other clusters.

2 types of clustering

- partitional clustering
- hierarchical clustering

03

Partitional Clustering

A set of clustering algorithms where each data point can only belong to one cluster.
Example: k-means algorithm

In k-means clustering, the goal is to divide the data into a predetermined value for K, the number of clusters. Each data point will fall into only one cluster of the K clusters, and therefore the clusters will not overlap like they would in hierarchical clustering.

04

Hierarchical Clustering

finds clusters by a system of hierarchies (clusters within clusters)

Every data points can belong to multiple clusters, some clusters will contain smaller clusters within it. This hierarchy system can be organized as a tree diagram.

2 ways of heirarchical clustering

- top-down (whole cluster to single points)
- bottom-up (single points to big clusters)

05

Agglomerative algorithms

will find clusters with a bottom-up approach

> Start with each data point as a cluster and progressively "zoom out" and combine smaller clusters into larger clusters.

Divisive algorithms

will find clusters with a top-down approach

> Divisive algorithms start out by looking at the entire dataset as one cluster, then "zooming in" to divide the dataset into smaller clusters.

Data-Driven Science

FOLLOW US FOR REGULAR UPDATES ON DATA SCIENCE

At DDS, we will help you build a career in Data Science

LINK IN BIO