Module 6: Data Clustering and Principal Component Analysis (PCA)

Glossary

DBSCAN

A density-based clustering non-parametric algorithm that groups together data points that are closely packed together, marking as outliers points that lie alone in low-density regions

K-Means

A method of vector quantization that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster

K-Means++

An algorithm for choosing the initial values or 'seeds' for the k-means clustering algorithm

Principal Component Analysis (PCA)

The process of computing principal components and using them to perform a change of basis on the data, sometimes using only the first few principal components and ignoring the rest

Singular Value Decomposition (SVD)

The factorization of a real or complex matrix