K NEAREST NEIGHBOR

K Nearest Neighbor Theory

KNN is a supervised machine learning algorithm for regression and classification problem. It uses data to classify new data points based on similarity measures (example: distance function). In KNN classification, a new data point is classified by the plurality vote of its neighbors, with the data point being assigned to the class most common among its *k* nearest neighbors (*k* is a positive integer).

K Nearest Neighbor with Python

Data Source: https://archive.ics.uci.edu/ml/datasets/Glass+Identification

This is a Glass Identification Data Set from UCI. The main objective of the analysis is to classify data points based on their glass type with K Nearest Neighbor Classifier.

Attribute Information:

1. Id number: 1 to 214 (removed from CSV file)
2. RI: refractive index
3. Na: Sodium (unit measurement: weight percent in corresponding oxide, as are attributes 4-10)
4. Mg: Magnesium
5. Al: Aluminum
6. Si: Silicon
7. K: Potassium
8. Ca: Calcium
9. Ba: Barium
10. Fe: Iron
11. Type of glass: (class attribute)
    -- 1 building windows float processed
    -- 2 building windows non-float processed
    -- 3 vehicle windows float processed
    -- 4 vehicle windows non-float processed (none in this database)
    -- 5 containers

-- 6 tableware
-- 7 headlamps