

Apply K-means algorithm to cluster a set of data stored in a .csv file.

Algorithm:-

- Step 1: Select the number K to decide the number of clusters
- Step 2: Select random K points or centroids.
- Step 3: Assign each data point to their closest centroid, which will form the predefined K clusters.
- Step 4: Calculate the variance and place a new centroid of each cluster.
- Step 5: Repeat the 3rd step, which means reassign each datapoint to the new closest centroid of each cluster.
- Step 6: If any reassignment occurs, then go to Step 4 else go to finish
- Step 7: The model is ready.

Apply EM algorithm to cluster a set of data stored in a .CSV file. Compare the results of K-means algorithm and EM algorithm

Algorithm :-

- Step 1 : Given a set of incomplete data, consider a set of starting parameters
- Step 2 : Expectation step (E-step): Using the observed available data of the dataset, estimate the values of the missing data.
- Step 3 : Maximization step (M-step): Complete data generated after the expectation (E) step is used in order to update the parameters
- Step 4 : Repeat step 2 and step 3 until convergence.

Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs.

Algorithm:-

Step 1: Read the given data sample to X and the curve (linear or non linear) to Y

Step 2: Set the value for Smoothing parameter or free parameter say τ

Step 3: Set the bias/Point of interest set x_0 which is a subset of X

Step 4: Determine the weight matrix using:

$$w(x, x_0) = e^{-\frac{(x - x_0)^2}{2\tau^2}}$$

Step 5: Determine the value of model term parameter β using

$$\hat{\beta}(x_0) = (X^T W X)^{-1} X^T W y$$

Step 6: Prediction = $x_0 * \beta$

22/06/22

LAB-6

Write a program to construct a Bayesian Network considering Training Data
Use this model to make predictions.

Algorithm →

1. First identify which are the main variables in the problem to solve. Each variable corresponds to a node of the network. It is important to choose number state for each variable.
2. Second, define structure of the network i.e. the causal relationship between all the variables.
3. Third, define the ~~problem~~ probability rules governing the relationships between the variables.

Program No: 8
KNN - K-Nearest Neighbour Algo.

→ Lazy Learning:

Suppose we have a new data point. which category do we assign it to. we use KNN for Euclidean distance (d)

Algorithm

$$d = \sqrt{(x_{c1} - x_{a1})^2 + (x_{c2} - x_{a2})^2}$$

1) Step 1: Select number of k neighbours

Step 2: Calculate euclidean distance of k no of neighbours

Step 3: Take k nearest neighbours as per Calculated Euclidean distance

Step 4: Among data k-neighbours, count the no of data points in each category

Step 5: Assign new data point to category for which no of neighbours is maximum

Step 6: END

Teacher's Signature : _____