

Revise :-

1. KNN as a Classification & Regression Algo.
2. Templates use for Classification & Reg^r Algo.
3. Evaluation Metrics of Classification Algo.
(g) f1 score (a) Precision (d) Confusion Matrix
(b) Recall (e) Error Rate
(c) Accuracy (f) Classification Report
4. KNN on a Regression data.

Agenda :-

1. Regression Metrics in python using scikit learn.
2. K-Means Clustering Algorithm.
3. Implementation of K Mean in python using sklearn.

* Regression Evaluation Metrics

(1) Mean Absolute Error

(2) Mean Squared Error

(3) Root Mean Squared Error

(4) R^2 Error ✓ ★

(5) Adjusted R^2 Error. ✓ ★

} Very Imp

* Euclidean distance calc from Centroid Point to data points
 1st row
 $C3 :- (8, 4)$ $(x_1, y_1) :- (2, 4)$
 (x_2, y_2) (x_1, y_1)

ED :-

$$\sqrt{(8-2)^2 + (4-4)^2}$$

$$= \sqrt{6^2 + 0^2} = \sqrt{36} = \underline{\underline{6}}$$

$$C1 :- \begin{matrix} (x_1, y_1) \\ (2, 4) \end{matrix} \begin{matrix} (x_2, y_2) \\ (2, 6) \end{matrix} \begin{matrix} (x_3, y_3) \\ (4, 7) \end{matrix}$$

$$\underline{\underline{CP}} :- \underbrace{x_1 + x_2 + x_3 + \dots}_n, \underbrace{y_1 + y_2 + y_3 + \dots}_n$$

$$= \left(\frac{2+2+4}{3}, \frac{4+6+7}{3} \right)$$

$$C1 = (2.67, 5.67)$$

(CP)

Revision :-

1. Dataset looks in a Unsupervised ML Algo (Target is absent)
2. Solved a problem using K-Means Clustering.
- Def ine a K-Means Algorithm.
4. K in K-Means represent the number of clusters.
5. Implementation of K-Means in python.

Agenda:-

1. Implementation of K-Means using sklearn in python.
2. Ideal method to select k (no. of clusters)
3. Elbow method.
5. Data preprocessing techniques.

Elbow Method:-

$$\underline{K=1} \rightarrow \underline{C_1}$$

$$K=2 \rightarrow C_1, C_2$$

$$K=3 \rightarrow C_1, C_2, C_3$$

$$K=4 \rightarrow C_1, \dots, C_4$$

$$K=5 \rightarrow C_1, \dots, C_5$$

$$\underline{K=20} \rightarrow C_1, \dots, \underline{C_{20}}$$

WCSS :- Within Cluster
Sum of Squares

:- used for judging
the optimal K .

Since we have check for each value of k between 1 to 20,

① We have use a for loop

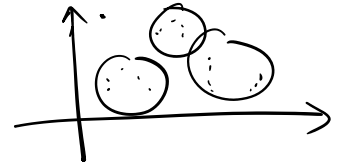
② $wcss = []$

③ loop running

calc wcss and that will be appended to

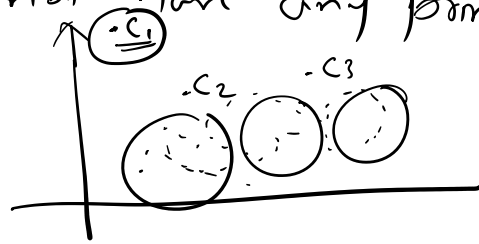
④ We create a visual ^{wcss empty list} to identify the optimal k .

init = K-means++



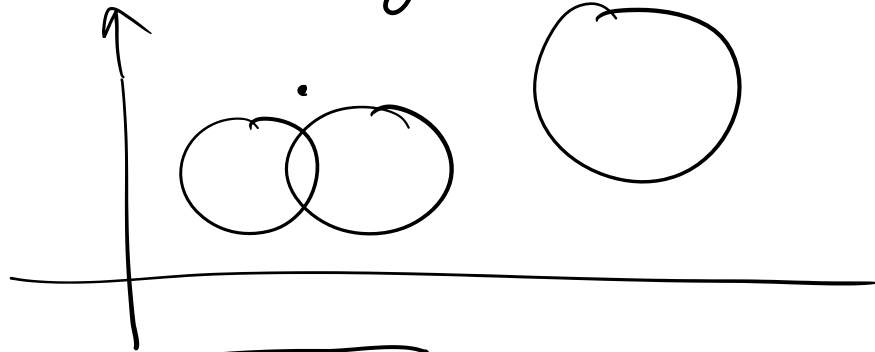
Standard KMeans Algorithm :-

1. Sensitive to the initialization of centroids or the mean points. If the centroid is very far, then that cluster will not have any point associated with it.



2. More than 1 cluster might end up linked with single same cluster.

Standard



init = 'K-means++'

K-means++

* K-means++ is the standard K-Means Algorithm coupled with smart initialization of centroid points.

* random_state = 42

* Simply used to make the predictions static

$$k=1$$

$$\boxed{\underline{\underline{133}}} \xrightarrow{15} \underline{\underline{132.5}} \text{ or } 133.5 \text{ or } \underline{\underline{134}}$$

ML Algorithms :-

1. You cannot supply a missing data
2. If any column (variable) is having higher values then ML Algo will give importance to that.
3. ML algo do not understand text data.
 eg:- E-112 E-^XName Cl[✓]val Age Salary Yrs. Exp. ^{Numerical}