**Unsupervised learning : unlabelled**

1. **Clustering**
2. **K-means algorithm (clustering)**

**We have 2 input in our dataset  and suppose we have 5 records/rows/**

**datapoints**

**x   y**

**1  12  39**

**2  20  36**

**3  28  30**

**4  18  52**

**5  29  54**

We have 2 features (input variables) X,Y and record=5

**Suppose there are 2 cluster K=2  C1 and C2**

**Suppose assign : first record in C1 cluster**

**Second record in C2 cluster**

**C1 C2**

**Means Xc=12,Yc=39 Means Xc=20,Yc=36**

**This is centroid of This is centroid of**

**C1 cluster C2 cluster**

|  |  |
| --- | --- |
| **Find the distance of**  **3rd record with centroid of C1 cluster Xc=12 and Yc=39.**  **The co-ordiate of 3rd record 28,30**  **Means x=28 and y=30**  **Find Euclidian Distance D =**  **Sqrt((Xc-x)2+(Yc-y)2)**  **sqrt((12-28)2+(39-30)2**  **sqrt((-16)2+(9)2**  **D=sqrt(256+81)=** 18.35 | **Find the distance of**  **3rd record with centroid of C2 cluster Xc=20 and Yc=36.**  **The co-ordiate of 3rd record 28,30**  **Means x=28 and y=30**  **Find Euclidian Distance D =**  **Sqrt((Xc-x)2+(Yc-y)2)**  **sqrt((20-28)2+(36-30)2**  **sqrt((-8)2+(6)2**  **D=sqrt(64+36)=** 10 |
|  |  |

**Shortest distance of 3rd record from C2=10**

**Note : select shortest distance means**

**Min distance(18.35,10)=10 means 3rd record goto cluster C2**

**C1 C2**

**Find the new centroid point of cluster C2: means find the average /mean= (20+28)/2 ,(36+30)/2 => 48/2 ,66/2**

**New centroid of cluster C2=24,33 but centroid of cluster C1 is previous means 12,39**

**And 1st record hold in C1 cluster**

**And 2nd and 3rd record hold C2 cluster**

**Again same process repeat**

**C1 cluster C2 cluster**

|  |  |
| --- | --- |
| **Find the distance of**  **4th record with centroid of C1 cluster Xc=12 and Yc=39.**  **The co-ordiate of 4th record 18,52**  **Means x=18 and y=52**  **Find Euclidian Distance D**  **Sqrt((Xc-x)2+(Yc-y)2)**  **sqrt((12-18)2+(39-52)2**  **sqrt((-6)2+(13)2**  **D=sqrt(36+169)=** 14.31 | **Find the distance of**  **4th record with centroid of C2 cluster Xc=24 and Yc=33.**  **The co-ordiate of 4th record 18,52**  **Means x=18 and y=52**  **Find Euclidian Distance D**  **Sqrt((Xc-x)2+(Yc-y)2)**  **sqrt((24-18)2+(33-52)2**  **sqrt((6)2+(-19)2**  **D=sqrt(36+361)=** 19.92 |

**Shortest distance of 4th record from C1=14.31**

**Note : select shortest distance means**

**Min distance(14.31,19.92)=14.31 means 4th record goto cluster C1**

**C2**

**C1**

**Find the new centroid point of cluster C1: means find the average /mean= (12+18)/2 ,(39+52)/2 => 30/2 ,91/2**

**New centroid of cluster C1=15,45.5 but centroid of cluster C2 is previous means 24,33 And 1st ,4th record hold in C1 cluster**

**And 2nd and 3rd record hold C2 cluster**

**Again same process repeat**

**5th record**

**Center point=(x1+x2+x3)/3 ,(y1+y2+y3)/3**