## 13/12/2022 Machine Learning Day 11: Clustering Machine Learning Supervised Uns uper vised Machine Learning Machine Learning A Un supervised Machine Learning: Unsupervised Machine Learning used machine learning algorithm to analyze and cluster implabelled datasets. These algorithms the need of human intervention. Difference Between Supervised and Unsupervised Learning => In superised machine Learning, input data is provided to the model along with the output. The goal of supervised learning is to thain the model so that it can predict the output when it is given new data. => In imsupervised lealning only input data is phovided to the model. The good of unsupervised learning is to find the sthucture and patterns from the input data. Unsupervised learning does not need any supervision. Instead, it finds patterns from the data My its own.

	Unsupervised Machine Learning Algorithm
	A P CENT AND PERSON OF THE PER
0	K- means -> K. mean ++
3	Hi Char chical Clust exing
3	Doscan clustering
	For equilibrium and administration to the contract of the cont
	Un Superised machine Learning
	Server (1) = 1
	Height weight BMI Country
	170 60 21 FND
The second	180 65 22 UK
	160 TO 20 USA
	165 75 18 IND
	140 55 19 USA
	the 1st energy 155 757 The Mitches are
	In superised machine Leasuing we predict the
	target feature (BMI) using independent features
	(Height, weight).
	The same of the sa
	In Unsupervised machine Learning we make chuters
	vased on country column. Clustering means
	geouping of data.
	considering Winking out Without 1921
0	K- means:
	Data -> Similagity -> Distance -
	Data > Similarity > Distance -  y Euclidean Distance <
	31
and established	Phytha gohas Theo sem.
C	Phythagokas Theogen. $P^2 = P^2 + B^2$
	and the second of the second o

$H(P_1,P_2) = \sqrt{P^2 + B^2}$	
H(P1,P2) = 1 P2 + B2	
(D) K- meantain -> Promonant +	
$D(P_1, P_2) = \sqrt{(x_2-y_1)^2 + (y_2-y_1)^2}$	
Distance bet"	
ρ, ρ2	
Marking to always Museling Leading	
Euclidean Distance	
$D(P_1, P_2) = \int (x_2 - x_1)^2 + (y_2 - y_1)^2$	
parmus) the base to be a second secon	
the land and the state of the s	
DATASET: Perform Clustering on this clataret	
the street of the self-or the	
Height weight	
185 ADV72 PI 22 ONI	
110 56	
168 60 60	
toget scales (EMI) wind fine 8 molecul 179 was	
182 72 (twow thous)	
188 77	
To Washer was been to 17 me use of he was a second	
man 160 miles had 70 miles was more house	
183 84	
180 88	
180 67	
167 76	
the sound of a Similarity > Distance of	
Important point of 12-means:	
	0
O Centhoid & In K- means K denotes the	06
2 Distance L no of Centroid	)
3 mean	

2	Another impostant points:
	-> ELBOW me twod. Italian and the world
	> wcss: within cluster sum of Equare.
836	-> Intercluster
	> Intha cluster.
	to minimum we willed get that paints in a
	For evaluation of clustering me that
->	Dunn Indesc.
<i>→</i>	Silhoute Score / Sil houte Coefficient.
	Continue Classes Exist (E. 3) sursing and I and
X	
	out the Centroid and that we will find handomly
مهندد	Intially we take two centhoids.
	The state of the s
	Here 1st record 185 72 } These points we are
	and 2nd be cold 170 56 S considering as centroid
	and 2nd be cold 170 56 & considering as centroid and around these centroid we are creating our
	cluster.
Jan	Comming track to out date of Now us colostole
	First Centroid (1 (2 2nd Centroid. (185,72) (170,56)
	(185,72) (170,56)
	Deine to a December 1 to a point of the party of the part
(03)	Centroid is the center value abound which we are
	Creating our clusters.
o distance	action of the series both of the series of t
	Here we have 2 centroid that means our K
4 (60-72	value in K. means is 2.
	Turn 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
	We can take one cluster also , but everything will
	he inside one cluster only
	3.0P 3 (a. 1) (

	Now we have one Two Centhoids C, and Cz.
	Now we will calculate the Fuelidean distance
	between each centroid and au other points.
	means we take one point and calculate its
	distance from (1 and (2 both. whose distance
	is minimum we will put that point in that
	Respective centhoid.
	Electrician Property of the Control
	For example. Suppose for 3rd point.
	Distance $(C_1, 3) = 5$
-find	Distance ((2,3) = 8
mdomly	out the Contraint and that eve will find a
0	Here 3rd point is near to Centhold C1, so it will
	go in cluster Ci
we ase	
Lant Raid	in questional C, 5,3 or hands but
343	(185,72)
	The state of the s
	Comming back to our dataset Now we calculate oul
· Vina	actuall distance.
	(185,78) = 1 or (119,58) and the
	Height Weight C, (85,72) (168,60)
2,5	0 185 72 (85,72) (168,60)
	(2) 170 56 C2
	3 168 60 3ªd point Enclidean distance.
X	9 179 × 68 Marks S and
	(a) $182$ $72$ $= \sqrt{(168-185)^2+(60-72)^2}$
	© 188 71
Min	Ø 180 · 71 = √ 433
	(S) 160 70 Mas stands and abjunt 440
	(C1, 3) = 20.8

13 16	Now Distance between (2 and point 3 (170,56) (168,60)
T A	(170, 56) (168, 60)
	New C, = 185+179, 72+65
	Pistance (C2,3) = \( (170-168)^2 + (56-60)^2
	01stance ((2,3) = 4.472
	So Distance of point 3 is lower to C2 then
dated Ci	So Distance of point 3 is lower to C2 then  Ci, therefore point 3 will belong to C2 cluster.
	Cond capable C2
今	Similarly Calculate distance of point 4 from (, and
(-12 P	(C2 - 3) ( av 2 21) - 1) ( av 2 3) - 2 4 - 1 - 1 - 1
	Similarly Calculate distance of point 4 from (1 and (2) (185,72) (2 = (170,56) point 4 = (179,68)
	Distance (1,4) = \( (185 - 179)^2 + (72-68)^2
	= 7.211
	Distance (C2,4)= \((170-179)^2+(56-68)^2
	Liver to the second of the sec
	15 17 man de la clarife de la
1.5	Point 4 will belong to cluster C1
	The state of the s
	Now C, with he make updated
	Now when we gold point 3 to centroid (2
	the centhoid (2 will be upolated.
	New C2 = 170+168, 56+60
	(2-881) = 23
	New (2 = (169, 58)
	Markey will hainging dame so: (169, 56)

Similarly when we add point 4 to centroid C, New C, = 185+179, 72+68 New (1 = (182, 70) So parton or or point 2 is hower to Co the => Now Calculate distance of point 5 from updated (, and update CZ point 5 = (182,72) (1 = (182,70) (2 = (169,58)Distance (1,15) = \ (182-182)2 + (70-72)3 Distance (C2,5) = (182-164)2+ (72-58)2 = 19.10 point 5 will add to centroid C. Now C, will be more updated Update  $C_1 = \begin{pmatrix} 182 + 182 & 70 + 72 \\ \hline 2 & & & \\ \hline \end{pmatrix}$  $C_1 = (182, 71)$ (2 will homain same: (169,58)

wess = & distance (di, (k) 2

di in Ci

where c is the cluster centroids.

di is the data point in lach cluster.

	For K= n means 2, 3, 4 etc centhoid
Tell Til	Con dom shorten grad 3
	WCSS = 5 ( & distance (di, CK)2)
most a	$WCSS = \underbrace{\begin{cases} \mathcal{L} & \text{distance (di, Ck)}^2 \\ \mathcal{L} & \text{di in Ci} \end{cases}}$
	signs to
	where C is the cluster centhoids and
	d is the data point in each cluster.
	the first of a factors between the party of the transfer of th
	So If we calculate value of wess when K=1 then
	So If we calculate value of wess when K=1 then that value will be greater then wess value with
ance.	Wilk = 2 hads as hot sometide as the south of the
- डंगर्स	Introducted the same Turka charter this
	wess, > wess, > wess,
	- District ( Control of 1 184 - 187 ( sel Par-18)
	wess, Value of wess when k=1
	wesse: value of wess when k=2
	wess, Value of wess when k=1  wess : value of wess when k=2  wess : value of wess when k=3
	District (Creek and Horsen Heller (TZ-18)
a data	So. out wess Vs 1K Graph will become.
	Hem and the cluster certaind with a cluster
	1 ELBOW CHART At some point, you will
9,4	find sundden change and
	3 Then there will be no
	3 change.
	Me and after 5 the
	1 2 3 4 5 6 7 K value of wess when
	K = 5,6,7 will be same.
	dim Charles the Ch
	where C 18the dustes contectors.
	The second secon

Ques	Now Comming to the Question what should be the
	Now Comming to the Question what should be the value of K?
	(2) Sillante Scare
Ans:	In our case the K value should be 5, because
Two Y	it is the point after which there is no change in
stes	the value of wess.
	distance
	-Chateu au fai agast
	when K=1 Chust er will be pig
	and so the value of wess
	when K=1  Chust ex will be pig  and so the value of wcss
-	x * c x
	the state of the s
	Tam.
	Dunn Judese 2 min distance (26, 20)
	when K=2
	* * * * * * * * * * * * * * * * * * *
mā.	* C1 x x
sollo de d	eliet is for to x50 cluster than surger
	Chuster will be smaller
	and so the value of wess will be smaller then
	Silhonette Score = Li - ai (=×)
	max (a.b.)
	When 14=3
	and the second of the second
	$\left\{\begin{array}{c} x^{+}c_{1} \\ x \end{array}\right\}$
	The state of the s
*	Sil hove the riose value sanger ream - to
tel other	Now clusters will be much smaller because same
	points will be divided into 3 dusters.

	For the management of 3, 4 and constanted
a.11-	How to validate duster ?
	Dunn Index (DI)
(2)	Silhoute Score
	Are In our rose the K value should be 5, has
0	Dunn Index is calculated as a hatio of the smallest
	inter-cluster distance to the largest intra-cluster
	distance.
	- Clustees are far apart
510	Shipe to half of the state of t
55.2M ,	Dunn Index = min (Inter cluster distance)
	mase (Intra cluster distance)
	- Clusters are compact
-	Dunn Index = min distance (21, 20)
	masc distance (yi, yj)
	Mayan a particular and the second sec
A	
(2)	
	object is to its own cluster (cohesion) compared to other
	elusters (separation).
1635	- Alleria set alter soone is tulet attacked and alice and
-	Silhouette Score = bi - gi
	max (a,b)
	Seal assist the
	where a: Intha-cluster distance
	b: Nearest - cluster distance
	Silhouette score value langes from -1 to 1.
314055	1: me ans clusters are well apart from each other
	and clearly distinguished.