432
4
= 8

1) Introduction

2016-12-05

Data Mining

Definition:

Non-trivial extraction of implicit, previously unknown and potentially useful information from data.

Access log
↓
i.p, timestamp, url, size, response

- collaborative filtering

Bonferrow's Principle:

"If you look in more places for an interesting patter then that the amount of you data can support, there might be false patterns detected."

Data & Attributes

- data is collecⁿ of data objects
  - structured
  - unstructured data
  - data objects → reflection of truth

Data attributes r of 4 types

(i) Nominal : $a = b$ or $a \neq b$ relations are defined.
Eg: $M \neq F$ or $M = M$ but $M+F$, $M-F$ not defined
Eg: 1st name, last name, etc.

(ii) Ordinal : $=, \neq, >, <$
eg: divisions of exam, small, big.

Scanned by CamScanner

More randomness
= More information

iii) Interval : $=, \neq, >, <, +, -$
Eg:- age, length, volume, numeric ~~value~~ value, dates

(iv) Ratio : $=, \neq, >, <, +, -, *, /$
Eg:- temperature in Kelvin, monetary quantities, counts, age, mass, length etc

Statistical tools
Mean, Median, Mode, S.D(σ), Variance, Harmonic mean
Geometric mean.

① Nominal : Mode, Entropy
② Ordinal : Mode, Median, percentiles

Median is less prone to outliers than Mean.

③ Interval : Mode, Median, Mean, Standard Deviation (σ)
4) Ratio : M_o, M_d, $\bar{x}$, S.D, H.M, G.M

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N-1}}$$

Discrete & Continuous Attributes:

Scanned by CamScanner

Data
- → structural
- → Unstructural

Data Preprocessing

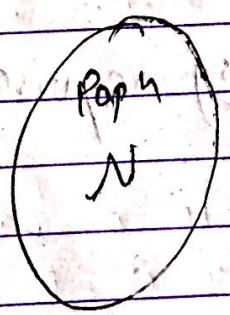
(i) Aggregation:

Dimentionality Reduction

2. Sampling simple
- ~~Single~~ Random sampling
- Sample without replacement
- Sampling with Replacement.
- Stratified Sampling

- Sample should possess similar POI (~~Point~~ property of interest as that of population

- For any given analysis, using sample will produce as good result as the original population.



→ sample 'n'.

Scanned by CamScanner

a * ←→ * b
Distance

Dimensionality Reduction:

- Dimensionality increases the sparceness of data
- Calculating distance & density ~~between~~ becomes compler

Curse of dimensionality

(i) Increasing dimensionality increase complexity to calculate distances

(ii) It increases sparceness

name	Age	Height

Purpose:-

- Reduce curse of dimensionality
- Reduce computational resources required to run an algorithm
- Allow easy data visualization
- Eliminate irrelevant features.
- → Reduce amount of time & memory required by data mining algorithms

Scanned by CamScanner

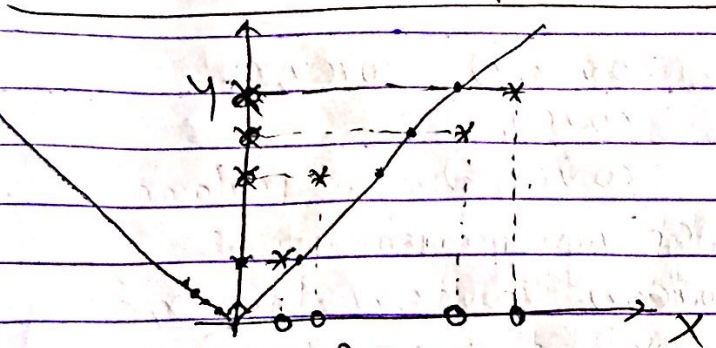
Assignment

Correlation.

① Covariance values: +, −, 0

(1) $x\uparrow$ & $y\downarrow$
(2) x & y are not related
(3) $x \propto y$.

Find the impact on the causality



PCA

x	y	→	x
			y

Doing so, amt of substantial information is lost.

Features: (Techniques)

(i) PCA : Principle Component Analysis
(ii) SVD => Single Value Decomposition

(iii) Feature subset selection
(iv) Feature creation

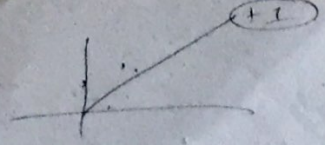
$$\Rightarrow \sigma^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$$

Co-variance

$$cov(x, y) = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

variation of x from y //

Scanned by CamScanner



Correlation is normalized form of covariance

$$\text{Corr.} = \frac{\text{cov}}{\bar{x}\,\bar{y}}$$

Principal Component Analysis:- (PCA)

Covariance Matrix

Appendices A & B

X
Y
Z

cov(X,Y) = cov(Y,X)
cov(Y,Z)
cov(X,Z)

Co-variance matrix.

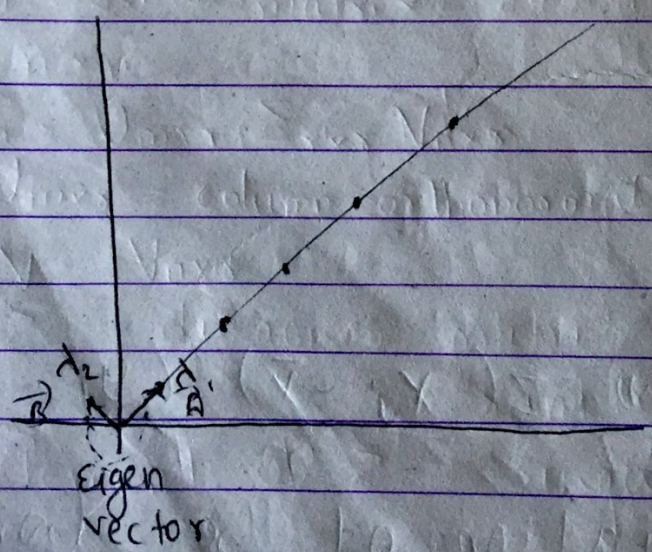
$$\begin{bmatrix} cov(X,X) & cov(X,Y) & cov(X,Z) \\ cov(Y,X) & cov(Y,Y) & cov(Y,Z) \\ cov(Z,X) & cov(Z,Y) & cov(Z,Z) \end{bmatrix}$$

Eigen vectors matrix

Eigen value:

$$A\vec{v} = \lambda\vec{v}$$

Eigen value.



$\lambda_1 > \lambda_2$ = signifies Eigen value of $\vec{A}$ will be greater than Eigen value of $\vec{B}$

Scanned by CamScanner

$$\begin{array}{r} 18 \\ 16 \\ \hline 34 \end{array}$$

$$\mathrm{cov}(X,Y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

X	Y		
1	2	-2	-4
2	5	-1	-1
3	6	0	0
4	7	1	1
5	10	2	4
$\bar{x} = 3$	$\bar{y} = 6$		

$\mathrm{cov}(X,Y) = \frac{18}{4} = \frac{18}{2} = 4.5$

$\mathrm{cov}(X,X) = \frac{10}{4} = 2.5$

$\mathrm{cov}(Y,Y) = \frac{34}{4} = \frac{17}{2} = 8.5$

$\mathrm{cov}(Y,X) = 4.5$

Co-variance matrix

$$\begin{bmatrix} 2.5 & 4.5 \\ 4.5 & 8.5 \end{bmatrix}$$

$r = 0.9762$

X'	Y'

$\det[A - \lambda I] = 0$

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 5 \\ 3 & 6 \\ 4 & 7 \\ 5 & 10 \end{pmatrix} \quad \bar{x} = [3 \;\; 6]$$

$$A = \begin{bmatrix} -2 & -4 \\ -1 & -1 \\ 0 & 0 \\ 1 & 1 \\ 2 & 4 \end{bmatrix} \quad A^T = \begin{bmatrix} -2 & -1 & 0 & 1 & 2 \\ -4 & -1 & 0 & 1 & 4 \end{bmatrix}$$

$$A A^T = \begin{bmatrix} -2 & -4 \\ -1 & -1 \\ 0 & 0 \\ 1 & 1 \\ 2 & 4 \end{bmatrix}_{5\times 2} \begin{bmatrix} -2 & -1 & 0 & 1 & 2 \\ -4 & -1 & 0 & 1 & 4 \end{bmatrix}_{2\times 5}$$

$$= \begin{bmatrix} 20 & 6 & 0 & -6 & -20 \\ 6 & 2 & 0 & -2 & -6 \\ 0 & 0 & 0 & 0 & 0 \\ -6 & -2 & 0 & 2 & 6 \\ -20 & -6 & 0 & 6 & 20 \end{bmatrix}$$

1	1.1	
1.5	1.6	
1.1	0.9	

x	
2.5	2.
0.5	0
2.2	2
1.9	2.
3.1	3.
2.3	2
2	1.

Scanned by CamScanner

Singular Value Decomposition : (SVD)

$$A_{mn} = U_{mm} \, S_{mn} \, V^T_{nn}$$

$$\boxed{U^T U = V^T V = I}$$

Columns of U are orthonormal eigen vectors of AA^T.
Columns of V are orthonormal eigen vectors of $A^T A$.

S is a diagonal matrix containing square root of Eigen values from U or V in descending order.

Assignment

Assume any matrix A. (Say $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$)
And perform SVD.
How to reduce dimensionality using SVD?

or, $A = U \Sigma V^T$

$U = U_{m \times m}$ = orthogonal matrix
Σ = diagonal matrix = $m \times n$
$V^T = n \times n$

Rotation
Stretch
rotate.

$M_{m \times n} = U_{m \times r} \, \Sigma_{r \times r} \, V^T_{n \times r}$

~~$M_{m \times n}$~~ $U_{m \times r}$ = column orthonormal matrix
$V = V_{n \times r}$
Σ = diagonal matrix
r = rank of M

Elements of Σ = singular values of M.

Scanned by CamScanner

Similarity / Dissimilarity:

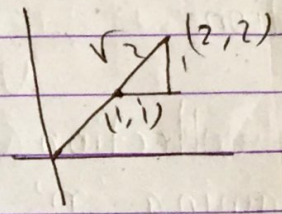
* Nominal:-

Dissimilarity

$$\text{Distance (D)} = \begin{cases} 0 & \text{for } p = q \\ 1 & \text{for } p \neq q \end{cases}$$

* Ordinal:

$$\left[\begin{array}{l} \text{short} \\ \text{medium} \\ \text{very large} \end{array}\right] \begin{array}{l} \} D = 1 \\ \} D = 2 \end{array} \Big\} D = 3$$



$$\sqrt{(1)^2 + (1)^2} = \sqrt{2}$$

* Interval & Ratio

(i) Euclidean Distance:

$$d(p,q) = \sqrt{\sum_{k=1}^{n} (p_k - q_k)^2}$$

(ii) Minkowski's Distance:

$$d(p,q) = \sqrt[r]{\sum_{k=1}^{n} (p_k - q_k)^r}$$

If $r = 1$, hamming dis (spell correc
If $r = 2$, Euclidean d

d(p, q)

(iii) ~~Mahdnobis~~ Mahalanobis Distance

$$S(p,q) = (p-q)\, \Sigma^{-1} (p-q)^T$$

Σ is the co-variance matrix o input data X.

Scanned by CamScanner

Properties of a distance metric

(i) $d(p,q) \geq 0$

(ii) $d(p,q) = d(q,p)$ for all p & q.

(iii) $d(p,r) \leq d(p,q) + d(q,r)$. [$\because$ Δ^{ar} $\Delta^{angular}$ inequality]

Chapter 3

Classification:

Given a collection of records (training set), with each record containing 'n' set of attributes and a class label, we need to

- find a model for class label as a function of other attributes.
- Records should be assigned a class as accurately as possible.

Use Test set for calculating the accuracy

- Supervised learning.

Scanned by CamScanner

Decision Tree Classifier
Rule Based "
Instance Based Classifier
Neural N/w "
Bayes Classifier "
SVM "

Neural Net.

Bayes $\Rightarrow$ $P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$

SVM = solve by increasing dimensionality

Decision Tree :-

A tree which predicts the value of class label (attribute) based on the values of known attribute.

Student Name	Age	Sex	Height	Division
R.S	21	M	---	? D
S.S	22	M	---	? F
M.S	25	M	---	? S
G.S	18	F	---	? F

Scanned by CamScanner

Hunt's Algorithm:

Let D_t be a set of training data set that reaches node 't'

If D_t contains records that belong to same class y_t → make it leaf node & assign y_t as class label

If D_t contains no records/is empty set → make it leaf node & assign default class.

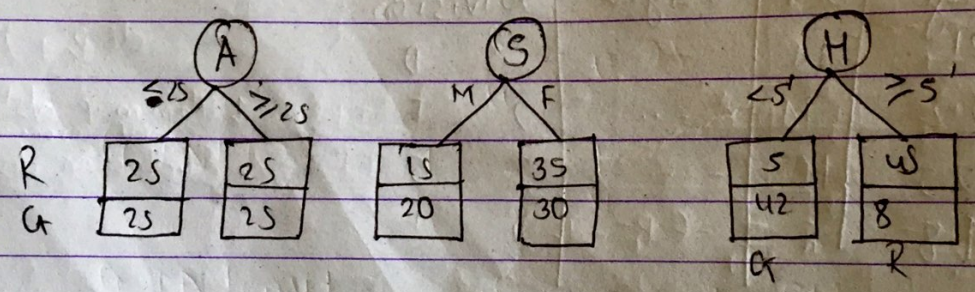
If D_t contains records with more than 1 class → Use one of the attribute test to branch & repeat.

Eg: 2.2

D [100]
Age (A)
Sex (S)
Height (H)

50 - Red (R)
50 - G.

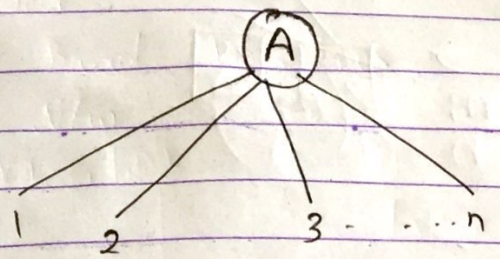
entropy ∝ heterogeneous.



Using Greedy Algorithm, 'H' is better bcoz

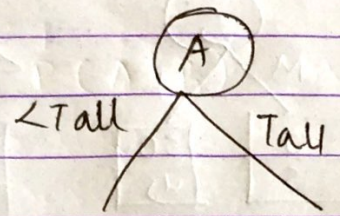
What are different attribute test conditions?

1st Name (Nominal)



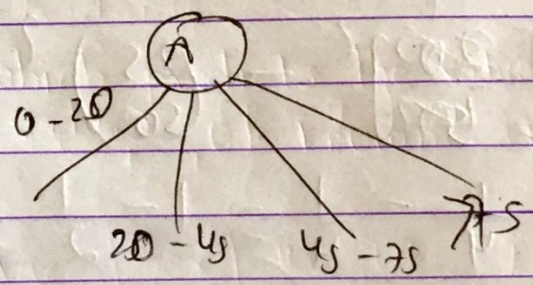
② For Ordinal.

Height



③ Interval & Ratio Attribute

Age



Equations to determine best split.

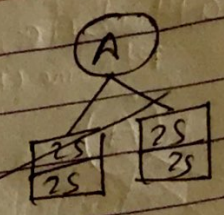
① <u>GINI Index</u>:

$$GINI(t) = 1 - \sum_j [P(j/t)]^2$$

$$GINI_{split} = \sum_{j=1}^{k} \frac{n_i}{n} GINI(i)$$

Scanned by CamScanner

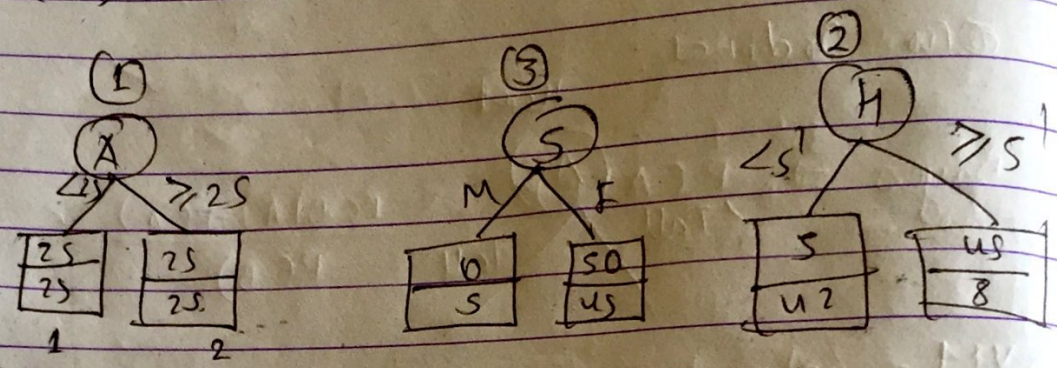
j = Class
P = Node



From Example 2.2
For A.

GINI( ) = 1 − Σ

Eg:



j = Solution

$$GINI(1) = 1 - \sum_j [P(j/t)]^2$$

$$= 1 - \left[\frac{\cancel{30}\,25}{\cancel{50}\,2}\right]^2 - \left(\frac{25}{50}\right)^2$$

$$= 1 - \frac{1}{4} - \frac{1}{4} = 1 - \frac{1}{2} = 0.5$$

$$= \cancel{\frac{3}{4} = 0.75}$$

$$GINI(2) = 1 - \cancel{\frac{1}{4} = 0.75}$$

$$GINI_{split} = \frac{50}{\cancel{100}\,2} * 0.5 + \frac{50}{\cancel{100}\,2} * 0.5$$

$$= \frac{1}{4} + \frac{1}{4}$$

$$= 0.5$$

For 2 (H) :

$$\cancel{GINI(1) = 1 - \left(\frac{47}{100}\right)^2 - \left(\frac{53}{100}\right)^2}$$

$$= 0.4982$$

$$GINI(1) = 1 - \left(\frac{5}{47}\right)^2 - \left(\frac{42}{47}\right)^2$$

$$= \cancel{0.2844}$$

$$= 0.19$$

$$GINI(2) = 1 - \left(\frac{45}{53}\right)^2 - \left(\frac{8}{53}\right)^2$$

$$= 0.26$$

Now

$$GINI_{split} = \frac{47}{100} \times 0.19 + \frac{53}{100} \times 0.26$$

$$= 0.23 \text{ (minimum)}$$

For, S :

$$GINI(1) = 1 - \left(\frac{0}{5}\right)^2 - \left(\frac{5}{5}\right)^2 = 0$$

$$GINI(2) = 1 - \left(\frac{50}{95}\right)^2 - \left(\frac{45}{95}\right)^2$$

$$= 0.4986$$

Now

$$GINI_{split} = \frac{5}{100} \times 0 + \frac{95}{100} \times 0.4986$$

$$= 0.4736$$

Scanned by CamScanner

degree $\propto$ no. of branches

To overcome drawbacks of GINI Split.

$$\text{Gain Ratio} = \frac{\text{Gain Split}}{\text{Split Info}}$$

$$\text{Split Info} = -\sum_{j=1}^{k} \frac{n_i}{n} \log\left(\frac{n_i}{n}\right)$$

Scanned by CamScanner

Entropy

$$\text{Entropy} = \sum_i P_i \log\left(\frac{1}{P_i}\right)$$

Entropy means randomness & corresponds information.

* Splitting based on Entropy:

$$GAIN_{split} = \text{Entropy}(P) - \left[\sum_{j=1}^{k} \frac{n_j}{n} \text{Entropy}(j) \right]$$

Overfitting: exactly fits in sample data but not in test data.

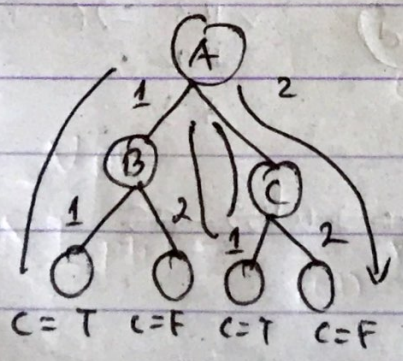
Advantage

(i) Transparency:

Model is transparent.

ii) Extremely fast (for sparse tree)

Rule based Classifier:-



If A=1 & B=1, then C=T

If A=1 & B=2, then C=F

If A=2 & C=1, then C=T

If A=2 & C=2, then C=F

Scanned by CamScanner

If p ; then q.

p → Antecedent

q → consequent.

Advantages

i) Better performance in some cases

(ii) Every route has same priority in decision tree but we can assign priority to each route in rule based

* Coverage:

* Accuracy

1) Mutualy Exclusive Rules

2) Exhaustive Rules:

Generalization

R1 ⇒ C

R2

R3 ⇒ C

R4

R5

data

(R1, R2, ... Rn)

If rules are not mutually exclusive, ~~arrange rules as~~

(i) assign priority to rules.

(ii) .

Scanned by CamScanner

H RIPPER Algorithm

1) Start from empty rule
2. Add conjuncts that maximizes FOIL Info gain.

$$Gain(R_0, R_1) = t\left[\log\left(\frac{p_1}{p_1+n_1}\right) - \log\left(\frac{p_0}{p_0+n_0}\right)\right]$$

t = no. of +ve data instances covered by both R_0 & R_1
p_n = no. of data that +vely match rule n.
n_n = " " " " -vely " " "

eneral to specific

ecific to general
If there are 100 data sets, 1st create 100 rules & generalize by adding conjuncts to combine rules.
RIPPER algo.

$R_0 = \{\ \} \Rightarrow C = T$

100	C = T ⇒ 50
	C = F ⇒ 50

Here, $p_0 = 50$, $n_0 = 50$

$R_1 = \underbrace{\{a < 2\}}_{70} \Rightarrow C = T$

$\longrightarrow$ C = T ⇒ 50
C = F ⇒ 20.

Here, $p_1 = 50$
$n_1 = 20$

p = matches antecedent & consequent
n = " " but not consequent

Termination for decision = Hunt's algorithm

10

Terminations algorithm for Rule Based Algorithm

CN2 / C4.5

General to Specific way to rule induc^n.

Instance Elimination:

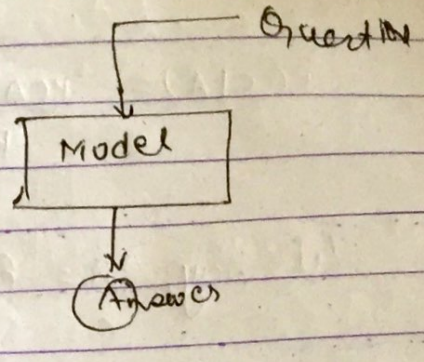
Suppose, Popul^n = 100

Scanned by CamScanner

Rajeev

Instance Based Classifiers:

- Supervised learning.



Here, training data itself is a model.

$d = 1 - s$

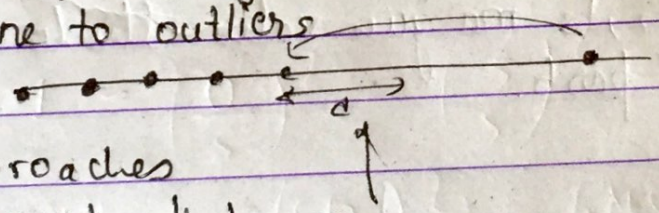
	Name	Age	Sex	Height	Class label (CL)
d_1	–	–	–	–	FD
d_2	–	–	–	–	SD
d_3	–	–	–	–	Fail
Q	–	–	–	–	→

1) Define distance metric

Disadv. of Instance Based Classifiers:

(i) Not suitable for large datasets.

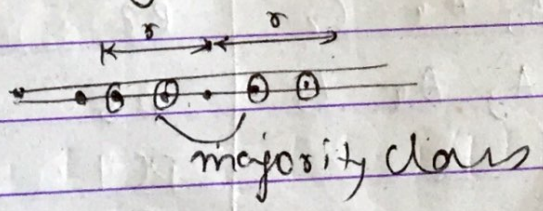
(ii) It is prone to outliers



Soln approaches

a) Threshold distance.

b) Majority voting with nearest neighbours.



c)

Scanned by CamScanner

Bayes Classifier:

$$P(C|A) = \frac{P(A|C)\,P(C)}{P(A)}$$

Let's say, A = Event of occurence of clouds
C = " " raining.

$$P(C_j|A_1A_2A_3A_4) = \frac{P(A_1A_2A_3A_4|C_j)\;P(C_j)}{P(A_1A_2A_3A_4)}$$

A_1 = Name A_4 = Height
A_2 = Age
A_3 = Sex

$C_j = C_1 = P1$
$= C_2 = P2$
$= C_3 \quad P3$

P_1, P_2, P_3 ma jun value ttulo hunb tehi assign garnu parb.

k	
$A_1\ A_2\ A_3\ A_4\ \ C1$	P1
$A_1\ A_2\ A_3\ A_4\ \ C_2$	P2
$A_1\ A_2\ A_3\ A_4\ \ C3$	P3
C1	PC1
C2	PC2
C3	PC3
$A_1\ A_2\ A_3\ A_4$	PA

Scanned by CamScanner

Naive Bayes :

Assumption :-

All the attributes are statistically independent.

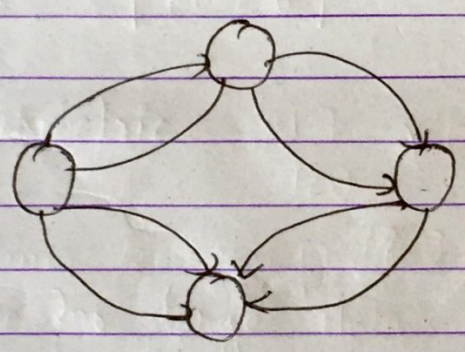
| N | A | S | H |

If the attributes are statistically independent.

$$P(A_1 A_2 A_3 A_4) = P(A_1) P(A_2) P(A_3) P(A_4)$$

$$P(C_j | A_1 A_2 \dots A_n) = \frac{P(C_j | A_1) P(C_j | A_2) \dots P(C_j | A_n) P(C_j)}{P(A_1) P(A_2) \dots P(A_n)}$$

Baysian Belief Network : (Study in detail @ home)



Assignment :- nominal

Sex	Height	Weight	Class
			→ Over weight → Normal → Underweight

Class label

BMI → formula

"Bayes classifier"

- Training data (S, H, Wt, BMI → formula)

Scanned by CamScanner

2017 - March - 6

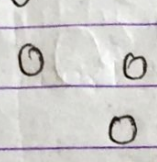
Clustering:-

Unsupervised Learning

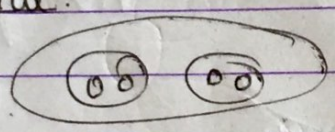
- group objects together such that objects in one group are similar to each other.
- Clustering groups data objects together such that objects that fall in a group are more similar to each other than the objects falling in different groups.
- no prediction
- generic grouping

Types of Clustering:-

(i) Partitional:

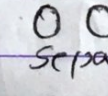


(ii) Hierarchial:



Types of Clusters:

(i) Exclusive vs Non-Exclusive



Separated

Distinct

(ii) Fuzzy vs Non fuzzy

In a fuzzy clustering, every object belongs to every cluster with a membership weight that is between 0 (~~absolutely~~ (absolutely doesn't belong) & 1 (absolu belongs.

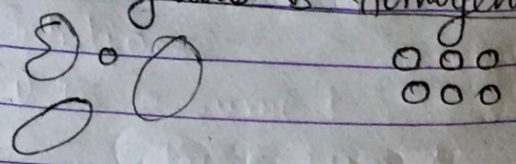
Dissimilarity: distance matrix : cartesian coordinate to visualise.

(3) Partial vs Complete

subset ko matra clusterip = Partial.

complete popn ko " = Complete

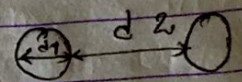
4) Heterogenous vs Homogenous



Types of clusters

1 Well separated Clustering:

d2 > d1

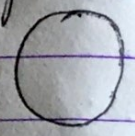


globular separated

2D = circular

3D = sphere

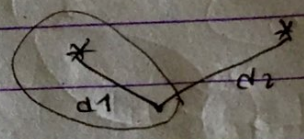
nD = globular



2. Center based clusters:

Centroid: central tendency of clusters

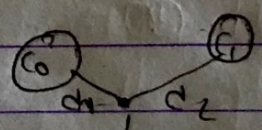
Here d1 < d2 so →



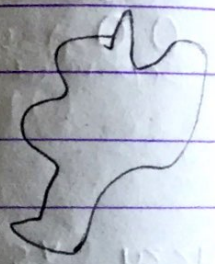
3. Contiguity based clusters

clusters ko centroid nabhaye nearest pt ma depend garne

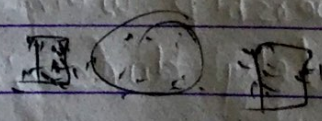
-MST



So,) falls in C0.



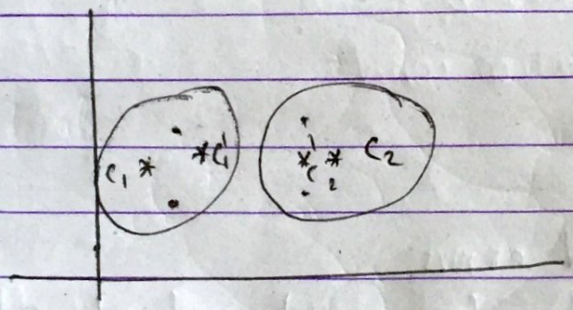
4) Density based clusters ✓ : any shape



(5) Objective funcⁿ clusters

K-means Clustering algorithm :- Generates partitional Clustering (Center-based clusters)

1) - Select 'k' points as initial centroid.
   ~~- From "k" clusters~~
2) - Form "k" clusters by assigning points to closest clusters
3) - Recompute the centroid
4) - Repeat until termination condition



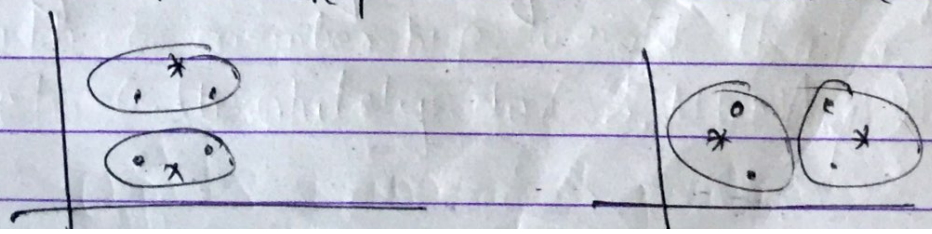
Termination conditions

i) Centroids donⁿt change
(ii) No. of iterations (when pts jumps betwⁿ clusters).

Limitations :

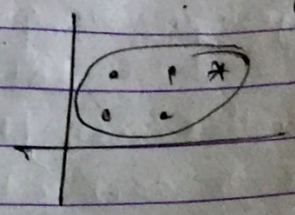
① We need to assume no. of clusters 'k' priorly.
i.e. How to assume 'k'?

② Final solⁿ is dependent of initial assumption



Scanned by CamScanner

(3) Null clusters are generated (*)



Solutions:

i) To assume k.

Bisecting k-means.

(ii) To ~~fix~~ solve soln dependability.
Use the whole algorithms multiple times and select the best assumption with lowest SSE with minimum SSE.

iii) Null cluster.

SSE = Sum of squared errors

Select the largest cluster and select the farthest pt from the centroid and assign the farthest pt as new centroid of a new cluster.

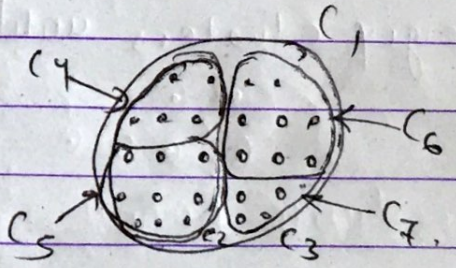
Scanned by CamScanner

II) # Bisecting K-means:-

Generates partitional & Hierarchial Clustering.

1. Initialize list of clusters to ~~certain~~ contain one cluster containing all data objects.
2. Select a cluster from list-of-clusters.
3. For no. of iterations do
   bisect selected cluster using k-means
   end for
4. Add the two clusters with minimum SSE to list of clusters
5. Until termination condition.

list-of-clusters = [$\cancel{C_1}$ $\cancel{C_2}$ $\cancel{C_3}$ C_4 C_5 C_6 C_7]

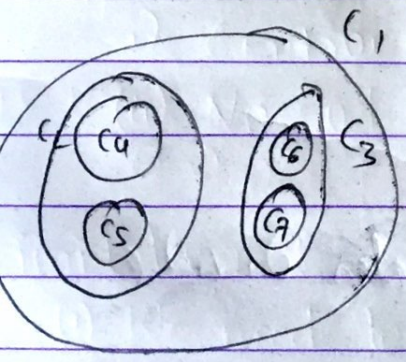


But

C4	C5
C6	C7

Partitional

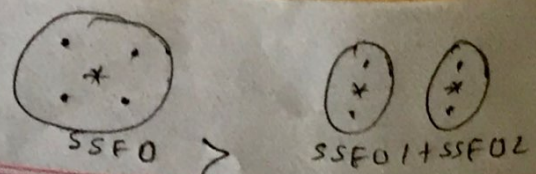
[C_1 C_2 C_3 C_4 C_5 C_6 C_7]



Hierarchial clusters.

Scanned by CamScanner

Gain = SSEO − (SSEO1 + SSE

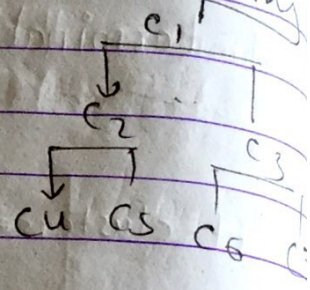


(3) Termination condⁿs :

4

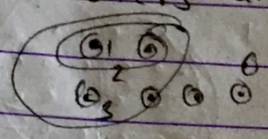
Hierarchial clusters

Dendogram



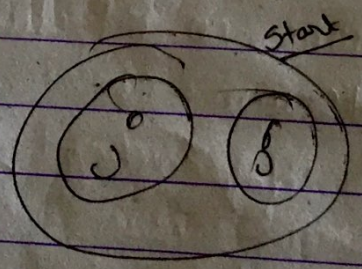
1) Agglomerative Clustering

Each pt is itself a cluster & we keep merging clusters until termination condⁿ



) Divisive Clustering :

All pts belong to one cluster, & we later divide the clusters



Eg :- Bisecting K-means Algo

III) Hierarchial clustering

Agglomerative Clustering: Proximity Based Algorithm.

(1) Compute the proximity matrix.

2. Let each data point be a cluster.

3. Repeat

Merge the two closest clusters.

Update the proximity matrix

Until Termination condⁿ.

Proximity matrix:

	P1	P2	P3	P4	P5
P1	0	1	2		
P2	1	0	3		
P3	2	3	0		
P4	1	1		0	
P5	1	2			0

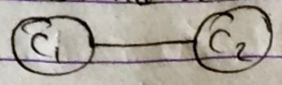
Merge P1 & P2

	P1 & P2	P3	P4	P5
P1&P2	0	1		3
P3	1	0		
P4	2		0	
P5	3			0

Termination condⁿ:

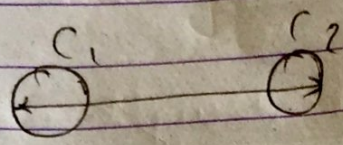
(H) Inter Cluster Distance

(1) MIN : distance betwn data objs in those clusters closest to each



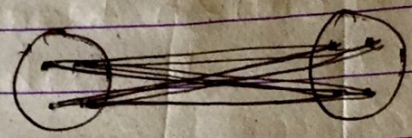
⇒ prone to outliers
⇒ non-elliptical shapes

(2) MAX :



: prone to outliers
⇒ Globular shaped

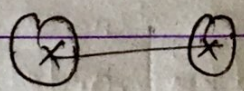
(3) Group Average



& Average

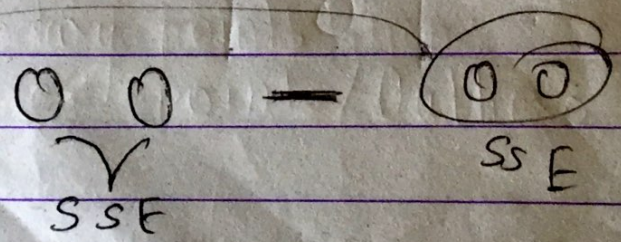
: Susceptible to outliers
⇒ Globular shaped

(4) Distance betwn centroids



⇒ Globular shaped

(5) Ward's method :



Scanned by CamScanner

Minimum Spanning Tree:

1. You need to join the nearest data objects with a line.
2. Break the longest line.
   Until terminationcond^n.

   - Generates contiguity based clusters.

Scanned by CamScanner

2017-03-09

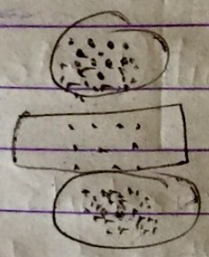
Data Mining: Introductory & Advanced Topics

Margaret H. DUNHAM

Chapter 1, 3, 4, 5, 6
+
Anomaly Detection.

Density Based Clustering:

Highly densed clusters are separated by from sparsely densed.



One of the algo. is;

I) DBSCAN:-

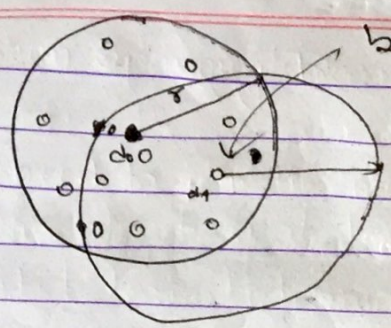
i) Core points : do
ii) Boundary points
iii) Noise points/Outliers

Given:

σ = Min^m no. of data objects

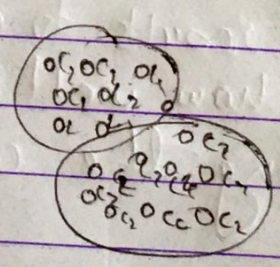
r = radius

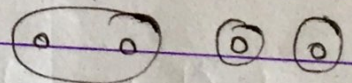
Scanned by CamScanner

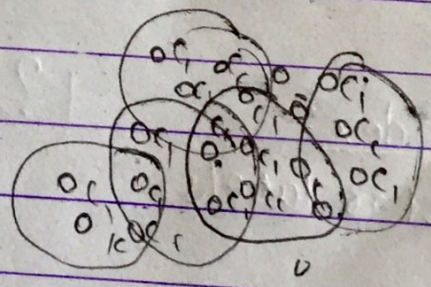


boundary pt - since it doesn't ~~cont~~ contain data obj $\geq \sigma$, but is a part of cluster

1. Initialize list of clusters
2. For every data object, repeat
   a) Check if its a core point
   b) If one-point, initialize a cluster lable & assign that to all other data objects.
   c) If only a boundry point, assign a cluster label of the one-point.
3. Until every data object.







Scanned by CamScanner

SSE_1

SSE_2

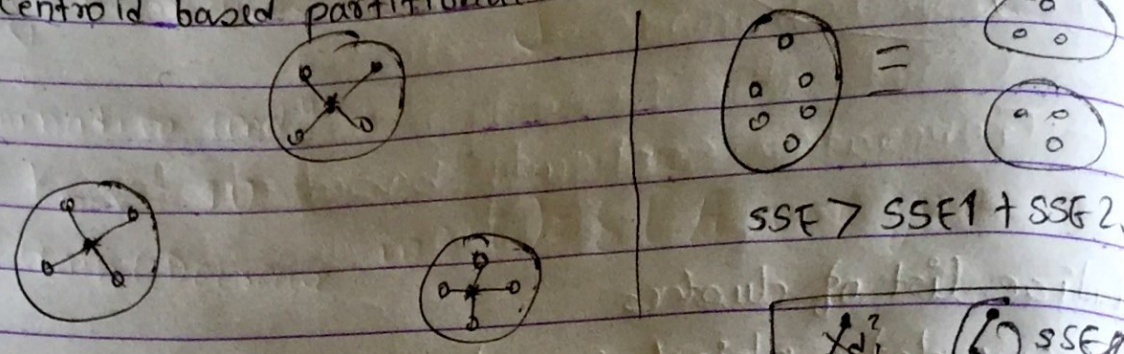
SSE_0

$SSE_0 > SSE_1 + SSE_2 + \delta$

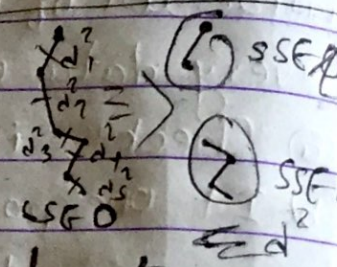
Cluster validity:

i) SSE = Sum of Squared Errors.

- used for centroid based partitional clusters.



$SSE > SSE1 + SSE2$



Threshold : δ

$SSE_0 < SSE_1 + \delta$ (???)

- can also be used for contiguity based clusters

Aspects of Cluster Validation:-

1. Determine the clustering tendency
2. Comparing results to ~~extremely~~ externally known results
3. Determine how well does cluster analysis fit the data
4. Determining the correct no. of clusters.

Measures of Cluster validation:-

1) External index: (?)

- cluster labels.

2. Internal Index
   - SSE
   - Correlation

3. Relative Index:

Cluster validation with correlation:-

1. Calculate two matrices
   a) Proximity matrix
   b) Incidence matrix.
2. Compute correlation between the two matrices
3. High value indicates nearby data objects are clustered.

Incidence matrix: same as proximity matrix but has only {0, 1} instead of distance

	P_1	P_2	P_3
P_1	1	0	
P_2			
P_3			

1 = same cluster
0 = different "

Higher correlation ⇒ better clustering

Scanned by CamScanner

Assignment:

- Assume 100 cartisian coordinate data objects.
- Cluster them to generate partitional clusters using ~~K-means~~ bisecting K-means algorithm.

- Termination cond^n
  (a) $K_{max} = 10$
  (b) Gain (Decrease in SSE) $= \delta$ [Assume your own δ