

Soft Computing

→ Approximation

If property applies to some members of a set, others in set are also assumed to have same properties.

Eg → If get up like politician, we assume politician decisions.

Inducing the properties of subset to the entire set with members having similar features.

→ Can't solve in analytical manner as NP problems.

We adapt from Nature (Nature is Optimizer)

Learning from Nature → Meta Heuristics
Heuristics of.

NP Complete/Hard v Reduced to P by guess & try

not done randomly but by heuristics of nature
This is called Meta Heur.

→ Overlapping Sets (can't say if member of set 1 or 2)

↳ Fuzzification : overlapping + Belongingness of member

in set is not 0/1
↳ features belongs to some extent

(defined by degree of membership)

→ We have set where we don't describe by exact numbers (like 2.5 for tall, 1 for short, etc) but like tall, very tall, short, etc.

If person is very tall, he ^{will} play volleyball well

not sure may be
5'9" or 5'11" or 6'2"

not exactly stated
meaning of wellness.

Such Relativistics in terms of linguistics. → soft.

These are dynamics of SC

Genetic Algorithms : adapted from genetics.

Charles Darwin → who survives & who dies?
→ which adapt to environment well.

Idea of Nature → More of particular species to have varied,
so that atleast some adapt.

Population → of sets of chromosomes.

Individual described by 23 chromosomes.

Set of Genes → Genotype

Expression of feature → ~~Geno~~ phenotype
(like dark eyes)

Each of us is a solution (based on chromosomes)

Problem → How to express for Eng. problems → using DS.

This bit stream represents one particular like array
SOLⁿ binary bit

* Chromosome \rightarrow Solution 101110101

Eg - if want to setup salon & there are candidates of many places, each bit represents 1 place
 0 → salon not set
 1 → salon set.

we make just 10 (instead of 2^{10}) solutions & give.

We ~~will~~ begin with smaller population as we are not doing exhaustive search.

* genes \rightarrow

(ADCTA) Alleles $\xrightarrow{\text{Group}}$ Genes

A group of ~~alleles~~ define ~~go~~ a feature (here 1 → for each place,

if multiple salons at a place, we can have 3 bits defining one place \therefore 3 positions define one feature

each allele is a posⁿ. Group of positions \rightarrow gene.
 group of genes define features.

Alleles $\xrightarrow{\text{group}}$ Gene $\xrightarrow{\text{group}}$ Feature

How to analyze a solution?

LIMSCIMATE
Date _____
Page _____

- No. of people & economic status → Profitability
- Get workers & cost of labour.

If Hotel, how far from farms?
~~far~~

Cost ↓ Profit ↑

We analyse using
Quality & Fitness

(Making some eqⁿ of these
bits)

↳ depends on problem how fitness is defined.

Eg -

$$f(x) = x + x^2 + \frac{\cos 42x}{2x+1} \quad 0 \leq x \leq \pi$$

Aim: To maximize/minimize this f^n .

We can solve by standard methods

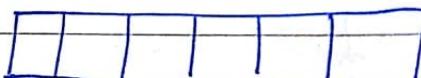
→ differentiation

→ approximation

→ Equating to 0

Other → Guess

Make a 64bit integer



This f^n is not fixed.
varies prob by prob.

TSP, space/time scheduling
Making cloud (Position)
Multi processor system

Robotics, Economics
Stock Price Prediction

Routing on chips → Connecting millions of components within chip.
with PDL path length

what to do with these 10 solutions?

evaluate these 10 & just chose best?

Cat & Tiger family same. Still cat survived, tiger not.
 Reason → Evolution Popⁿ by Popⁿ evolution happens.

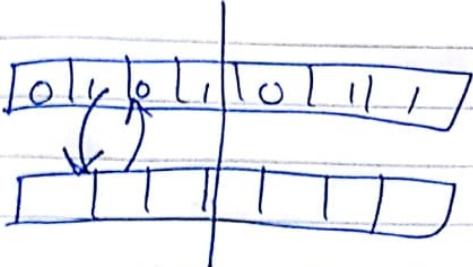
Reason of evolution in humans →

→ Reproduction (meiosis)

→ parts of chromosomes exchanged.

combⁿ of 2 chromosomes.

Eg →



one half of both sol's exchanged to form new solⁿ.

→ Crossover (similar to meiosis)

Better genes

Taking best of 2 individuals & mixing hoping to get best.



Exploitation & Exploration

Crossover → Exploitation → Building blocks → Mixing best

Mutation → Due to environment (radiation), genes get changed.

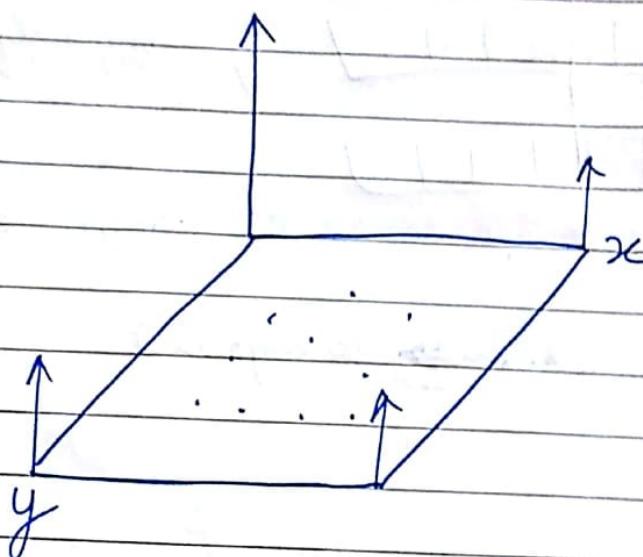
for this, we take random bit in chromosome & flip it

We are trying to explore a space for best
Feature Space

Parents \rightarrow Swimmer: Run x , Swimmer: Run x

Offspring \rightarrow Expected \rightarrow Swim ✓ Run ✓

Mutation can enable to climb tree also

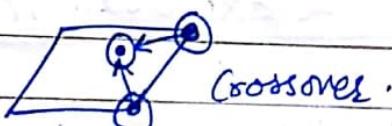


Point in
 $n-4 \rightarrow$
 Solutions.
 Other dim \rightarrow
 fitness.

Notion of

Distance b/w 2 solutions.

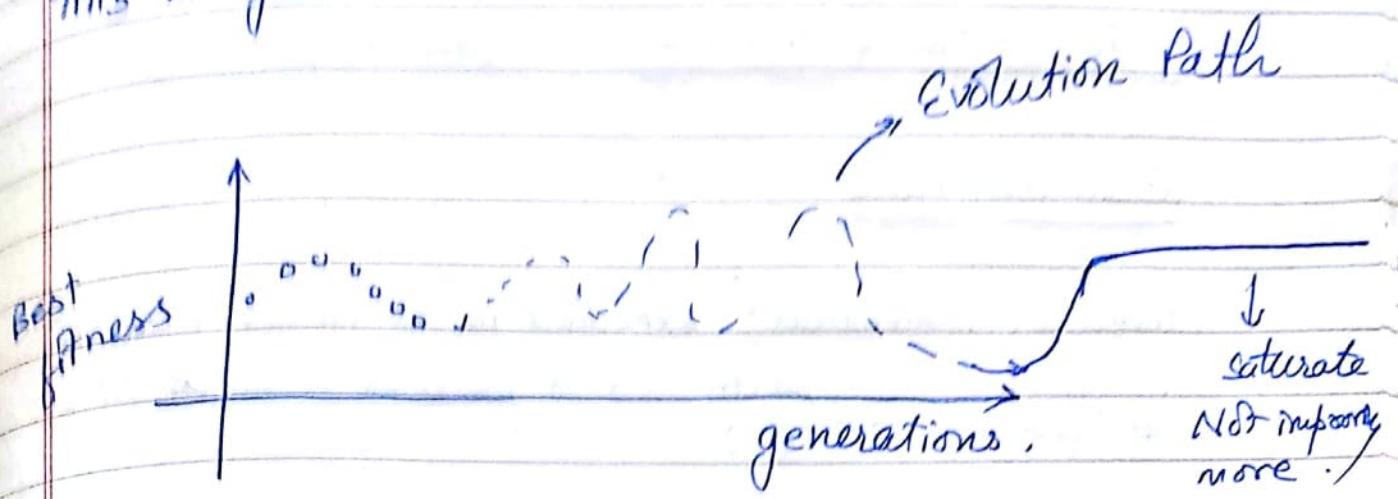
Nature \rightarrow move from 1 solution to another through crossover & mutation.



trying to search space through C & M.

Crossover & Mutation \rightarrow Genetic Operators
plus \rightarrow Crossing, Inversion,

this is guided search



we don't know if minima is global/local.

. We explore by C&M to find global min (of minimizer problem)

We have to stop somewhere as we can't explore the whole large space

where we stop may not be global.

\therefore soft computing doesn't guarantee global optimum.

Simple, Robust heuristic from nature but not complete.

Selection

from current popⁿ, whom to chose to generate next popⁿ

This is called Selection

We can combine strong with strong

but we can't reject weak ones also (can survive acc. to environment)

∴ selection done by -

Fitness Proportionate (Acc. to fitness)

Roulette time

Given an individual's expected no. of times it is allowed to reproduce (participate in process) is \propto fitness of i divided by sum of fitnesses

~~Exp-val (i)~~

$$\text{Exp-val}(i, t) = \frac{\text{fit}(i, t)}{\sum_{j=1}^N \text{fit}(j, t)}$$

We take Roulette wheel with sizes proportional to fitness & spin $\frac{N}{\text{pop size}}$ times & where ptr stops is allowed to participate.

Implementation

Sum of fitness $\rightarrow S$.

Those random no. $\leq S$

Add fitness of individuals. As soon it exceeds S, take that individual.

SC

Selection

If Population too homogeneous & less variety

Exploration Halt

Stuck at local Minima

can be in Roulette wheel (becoming many times same)

Normally, mutation v. less ($\sim 0.1\%$)

→ Macro Mutation

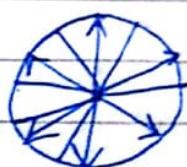
→ Increase Mutation

Chromosome Encoding → Creation of solutions (giving values)

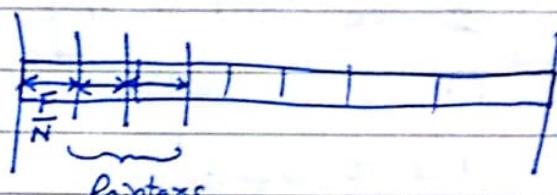
If length \uparrow , probability of each will \downarrow . will increase evolution process.

Stochastic Uniform Sampling

Rotate wheel only once. We have equi spaced pointers.



start
selects = rand()



$\text{ptr} := \text{start} + i * p \mid i \in I \rightarrow N$ $P = \frac{F}{N}$ → Total fitness

for ($i=0$; $i < N$; $i++$)

 for ($\text{sum} += \text{fit}(i, t)$, $\text{sum} < \text{ptr}$; ~~$\text{ptr}++$~~)
 select i ;

Initially, Selection Pressure ↑ ~~chance~~ given to lesser →

Selection Pressure more constant here.

Here, using CDF.

Instead if we take PDF, we work b/w 0 & 1
Guarantee → Prob. b/w.

?

$$[f_{\text{fitness}}] \rightarrow [f_{\text{fitness}}]$$

divide fit by total fit

Sigma Scaling

If fitnesses → Prob. Dist.

Initially, SD ↑

Convergence reduces SD.

fitter individuals come, SD ↓

Rate of evolution depends on SD

We want as SD ↓, chance of less fit still remains ~~somewhat~~ somewhat & is not 0.

$$\text{Exp-val}(i, t) = i + \frac{\text{fit}(i, t) - \bar{\text{fit}}(i, t)}{2\sigma}$$

As we want to stop when Mean too high, but it should not be too early. ∵ We are trying to explore more due to SD.

Selection Pressure becoming uniform.

Still linear proportionality

Simulated Annealing → Another Meta Heuristic

↳ Heat glass & gradually cool so that molecules align.

Boltzmann Selection

$$\text{Exp-val}(i|t) = \frac{e^{\text{fit}(i|t)/T}}{\langle e^{\text{fit}(j|t)/T} \rangle_t}$$

$\langle \rangle_t$
for a
time/
generation.

At high temp, diff. b/w expected value of most fit & least fit individual ~~decreases~~ increases with temperature

At high temp, most & least fit will have almost same prob.

This gives us control over Selection Pressure through T

Sigma → Low Pressure
SUS/Roulette → High Pressure

High Sel ⁿ Pressure	→ Exploitation
Low " "	→ Exploration

we start with high T & gradually decrease it to increase probability of highly fit individuals.

Complexity now polynomial \rightarrow No. of people in gen
 \times No. of gen.

We also like to decrease computation for every individual.

We want absolute computation also to be as less as possible & not only complexity



Real Time Applications

Parameters need to be changed acc. to environment to optimize (eg. angle of solar panel of rover, etc.)

In RTA & optimization problems of parameters, Meta-Heuristics is not much followed.

Make exp value proportional to ranks instead of absolute fitness.

Sort acc. to fitness & take ranks starting from 1

$$Exp = \text{Min} + (\text{Max} - \text{Min}) \times \left(\frac{\text{Rank}(i,t) - 1}{N-1} \right)$$

gradually, popⁿ concentrate on particular ~~indiv~~ species & ranks become same decreasing Max value & ∴ diff.

Pressure Constant

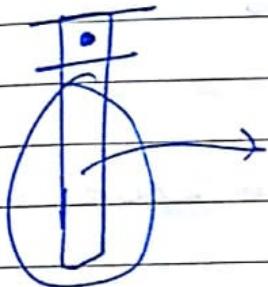
Eliticism

Few best individuals are transferred directly to next popⁿ.

Max fitness curve becomes non-decreasing.

Non-decreasing Best fitness Fⁿ.

Most Popular GA Technique



Tournament Selection

Take 2 individuals randomly.

generate rand no R & chose threshold value K say (.75)

If $x < K$ chose fitter one otherwise chose less fitter one & return both → do this N times

Use a library for
GA & play with it.
(Roulette) CLASSMATE

Date _____
Page _____

Tour + Elit

Elit + Rank

Elit + steady state

Steady state

In crossover, Best fit will be crossed with least fit chromosome.

Elitism



→ Don't crossover as similar
Do only mutation.

CO → Exploitation . But doing exploration to some extent by crossing with less fit.

DY Schema Theory → Proves why GA works.

→ Make Objective F^n

Double Pt. C^{oo}
Haplotype
deployed

classmate

Date _____
Page _____

Book

Göteborg

Read papers.

→ Feature Space (know features)

Real Nos.

→ Value of each feature

→ Constraints

→ Encoding Chromosome

→ Keep evolving

Convergence when best fitness becomes ~~not~~ constant.

Parameters

Crossover Rate { usually by experimental tuning.
Mutation Rate

or Make these part of chromosome.

	CR	MR
--	----	----

So that these rates also evolve.

Meta GA

~~Alley~~

1) Features controlling solution

Meta Heuristics when v. large space

Multi Dimensional Problems.

Eg →

Document analysis →

Comparing.
Sentiments
opinions

Topic Modelling
Summarizing.

Eg → Time Table Problem

Assignment Problem → Assigning Subject to Teacher

Scheduling Problem → Time Slots to classes

(Mapping in time dimension)

Knowledge

DB

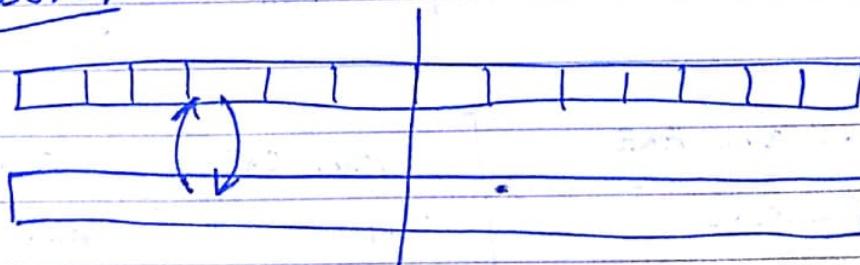
Eg → Specializations
Syllabus
Parameters.

If no selection & elitism

1 best will continue to next gen & ~~apply~~ apply

Genetic operators on rest population

Crossovers



Mutation

Randomly change a value.

Complexity = Population size \times No of Generations
 \times Complexity of Genetic Operators.

Parameters to be decided \rightarrow Mutation Rate

(Macro Mutation) \rightarrow Large Mutation Rate

~~Chromosome Encoding~~

~~Parameters & Selection Operator~~

~~Decide Domains~~

Chromosome Encoding

Decide Domains

constraints

Parameters

Fitness function & objective function

One way in TT \rightarrow Teacher satisfaction \rightarrow decide Levels for satisfaction

OF = TS Teacher Satisfaction

Distribution of subjects (spread)

(Pedagogical considerations)

Minimize Adj. b/w periods.

Room Closeness if space scheduling
 Also.

Give weights to each OF. can add max to num & min to denom.

$$\frac{w_1 OF_1 + \dots + w_3 OF_3}{w_2 OF_2 + \dots}$$

denom.

OF → Quality
→ Cost

either subtract / Divide

$$\frac{w_2 Q}{w_3 + w_1 C}$$

weights →
normalized
weights

if Probabilities, multiply.

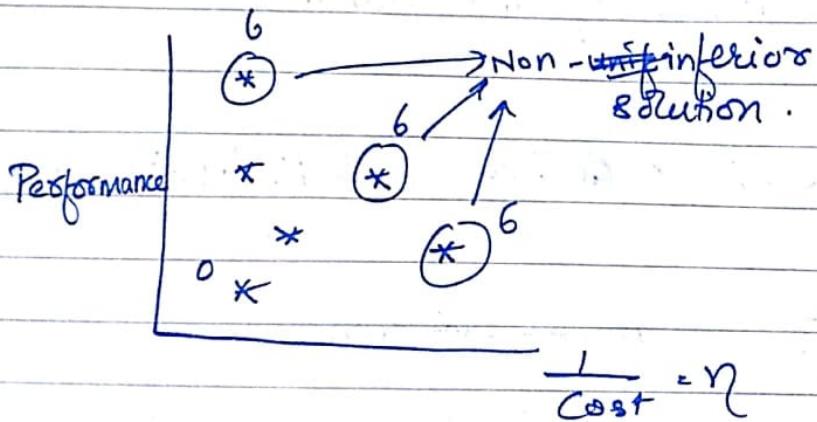
Selection operators guided by fitness

Genetic operators only on Solution Space, not on fitness.

GA → best for many non-differentiable eq'n's.

→ separation of concerns.

Suggestion



Rank → No of sol'n not superior to it in all aspects

To propagate all non-inferior solutions to next generation.

Pareto-optimal set

Pareto-optimality

Ant Colony Optimization SC
 Heuristic Pheromone
 η γ

X, C, f Combinatorial optimization problem.
 $\{x_1, x_2, \dots\}$ ~~Real~~ Discrete Space
 ↳ features
 $v_i \in D$ Continuous Space.
 v_1, v_2, \dots

ACO → More suited for combinatorial ~~Real~~ Problems & graph based problems

C → constraints.

$f: X \rightarrow R_0^+$ → set of positive real nos.

$x \rightarrow S \rightarrow$ solution.

$S_c \rightarrow$ feasible solutions in S .

$S, C, f: X \rightarrow R_0^+$

$f(S^*) \leq f(S) \forall S \in S_c$

↳ Global Optima (Minima Here)

ACO → Good for Graph Based (e.g.- TSP)

Pheromone → Deposited by ants on their path & other ants smells it to follow.

Set params, Init Pheromones.

spread Pheromones initially randomly

Particle optimization

Stener
Sethuraman Date _____
classmate Page _____

Set params, Init pheromones

Begin

Construct Ant Solⁿ → Search

Ants Synchronization - optional

Update Pheromone

End

stop after n iterations.

Ants may need to synchronize their paths.

sp → part of feasible solⁿ

$\alpha \rightarrow \text{A parameter}$
 $L \rightarrow \text{length of path}$
 $\eta = \frac{\alpha}{L}$

pheromone → Experience.

~~Heuristic~~ Heuristic (η) → Local knowledge.

$\rho = \text{Evaporation Rate}$

$$\tau_{ij} \leftarrow (1-\rho) \tau_{ij} + \rho \sum F(s)$$

summation for all ants travelled or max of all.
All ants incurrent path or only max travelled now.

$$S = S_{\text{iter}} \cup S_{\text{best so far}}$$

↓
Best in this iteration.

↓

Elitism : Best from starting in all iterations

ACO → Simpler than GA

→ used for Graph Based Problems.

→ Learning path optimization.

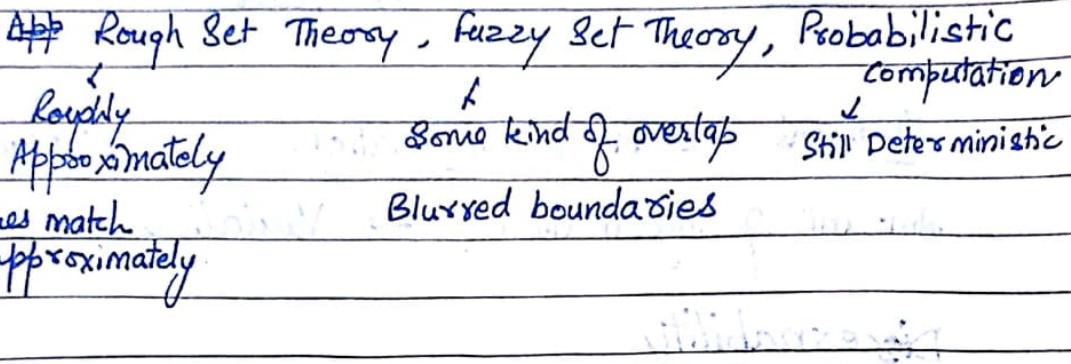
→ converges a bit faster → somewhat more exploitation



→ used for Combinatorial Problem in

SCRough Set Theory (Approximation)

Not same as
Probabilistic
Approximate

Supervised

→ Rule based Supervised Classification Technique

Say Database is -

Universal →

<u>U</u>	<u>Age</u>	<u>Height</u>	<u>Play Basketball</u>	<u>decision</u>
x_1	0-10	3-3.5	N	
x_2	0-10	3.5-4	Y	
x_3	10-20	4-6	Y	
x_4	10-20	4-6	N	
x_5	20-30	5.5-6.5	Y	← Acts like training data of Decision present apriority.
x_6	20-30	5.5-6.5	Y	
x_7	30-40	5-7	Y	

Should be able to classify & derive rules from the DB.

Information decision
system system

if () then, decision Y/N

Premium

$3-3.5, 3.5-4 \rightarrow$ Confidently N & Y
 \hookrightarrow Sure Rule.

10-20, 4-6 \rightarrow Possibility Rule

Shannon's Theorem of Information

$2^i \rightarrow$ Information

Amount of Information? \rightarrow Variation

Discernability

why reduce rules? Computation Complexity ↓
 Space ↓ Transfer Time ↓

How?

Divide into equivalence classes

Feature Reduction

if extent of discernability does not change on reducing features
 \rightarrow Shorter Rules.

B G A

Given a subset of attributes, how to define indiscernability?

Consider only age \rightarrow Equivalence classes will be cardinality

$$\text{IND}(\text{Age}) = [X]_{\text{Age}} = \{\{x_1, x_2\}, \{x_3, x_4\}, \{(x_5, x_6)\} | x_i \neq x_j\} = 4$$

$$\text{IND}(\text{Height}) = [X]_{\text{Height}} = 5$$

$$\text{IND}(\text{Age, Height}) =$$

If whole $\rightarrow 5$

If decision variable also taken $\rightarrow 6$ cardinality

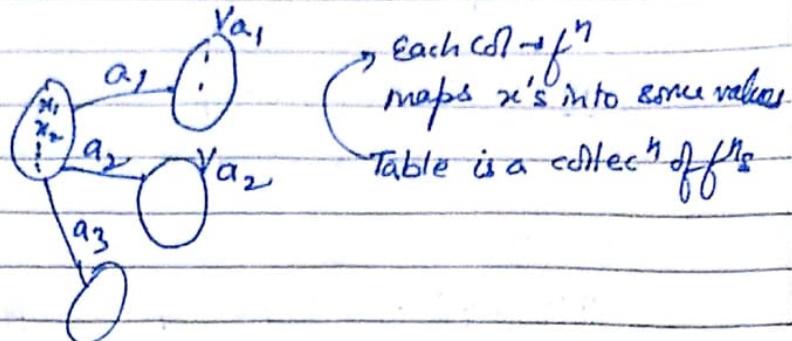
Premium

Completely Construction

Indiscernability of entire relational database for features B

 $\text{IND}_R(B)$

U	a_1	a_2	a_3
x_1	$v_{a_1}(x_1)$		
x_2		$v_{a_2}(x_2)$	
x_3			$v_{a_3}(x_3)$



$$\begin{aligned} & \{x : \forall x \in X : \\ & \quad \exists \{x_1, x_2\} \in U^2 : \\ & \quad \{x_1, x_2\} \in G_U : \\ & \quad \{x_1, x_2\} : \forall a \in A : v_a(x_1) = v_a(x_2)\} \end{aligned}$$

$$\text{IND}_R(B) = \{(x, x') \in U^2 \mid \forall a \in B : v_a(x) = v_a(x')\}$$

equivalence class of x w.r.t $B \rightarrow$ $[x]$

$$[x_1]_{Age} = \{x_1, x_2\}$$

Why choose many subsets? \rightarrow To get short rules without losing information.If all subsets \rightarrow exponential complexity \rightarrow v. large.By reducing features, infoⁿ lost but more support for classification
 \therefore accuracy ↑Approximation ↑ Degree of classⁿ ↓ Computation ↓ Premium

We try to approximate features to decisions.

Some classes in only Y or N but some can be in both. Where to put it?

X_Y : Set of all objects with decision X as Yes

$$B \subseteq A$$

$$\underline{B}X_Y = \{x \mid [x]_B \subseteq X_Y\}$$

We call it lower approximation if $\underline{B}X_Y \subseteq X_Y$

For Age,

$$\overline{\text{Age}}(X_Y) = \{x \mid [x]_B \cap X_{\text{play}} \neq \emptyset\}$$

$\{x_2, x_7, x_5, x_6, x_3, x_4\}$

Not in both

upper & lower Both

$$x_1$$

$$x_2, x_3, x_4, x_5, x_6, x_7$$

$$\overline{B}(X_C) = \overline{B}(X_C) - \underline{B}(X_C) \rightarrow \text{Boundary. } (x_3, x_4)$$

$$\overline{\text{Age}} = \{x_2, x_7, x_5, x_6\}$$

outside Region
 $U = \overline{B}$

$$\overline{\text{Age}}(X_{\text{play}}) = x_5, x_6, x_7$$

$$\overline{\text{Age}}(X_{\text{play}}) = \{x_1, x_7\}$$

$$\overline{\text{Age}} = \{x_1, x_2, x_3, x_4\}$$

Upper Possibility + Sure
Lock

Surely

outside Region = $\emptyset \rightarrow$ Externally undefinable.

$\overline{B} = B \Rightarrow$ No Boundary option \rightarrow all sure rules.

\Rightarrow Crisp Decision.

$$\begin{array}{l} \overline{B} \neq U \\ \underline{B} \neq \emptyset \end{array} \Rightarrow \text{Boundary}$$

Reducing features \Rightarrow More approximation More Rough More Support
Increasing \rightarrow less approx. \rightarrow But less support

Not much overfitting & not much rough. Premium
↳ only that instance. Working outside

$$\text{Accuracy } \alpha = \frac{|B|}{|\bar{B}|} = \frac{|B(x_c)|}{|\bar{B}(x_c)|}$$

None features \rightarrow May & classification can give false positives.

Finding no. of features to give good classification \rightarrow optimization problem.

Crisp $\rightarrow \underline{B} = B$

Externally Undefinable $\rightarrow \underline{B} = U$
 $\underline{B} \neq \emptyset$

Rough $\rightarrow \underline{B} \neq U$
 $\underline{B} \neq \emptyset$

Internally Undefinable $\rightarrow \underline{B} \neq U$
 $\underline{B} = \emptyset$

Totally Undefinable $\rightarrow \underline{B} = U$
 $\underline{B} = \emptyset$

All in Boundary Region.

To find optimal set of features \rightarrow
 if less features & less data \rightarrow Analytical Method
 if many features & more data \rightarrow Meta Heuristics.

Ques

$$IND(DEFR) = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$$

$$\underline{DEFR}(X_{accept}) = \{x_1, x_4, x_6, x_7\}$$

$$\overline{DEFR}(X_{reject}) = \{x_2, x_3, x_5, x_8\}$$

$$\underline{DEFR}(X_{reject}) = \overline{DEFR}(X_{reject}) = \{x_2, x_3, x_5, x_8\}$$

If ER \rightarrow still same data

If rough \rightarrow More sense of reducing ind. as may be good to be rough

$c_{ij} \rightarrow$ collection of all those attributes s.t. $x_i(a) \neq x_j(a)$

$$c_{ij} = \{a \in A \mid x_i(a) \neq x_j(a)\}$$

~~Rule based classification~~

~~Supervised~~: ~~we know some labels of training data already~~ {
For each pair, find diff. features & or them.
AND for each pair.

CLASSMATE

Date _____
Page _____

~~Supervised~~

	x_1	\dots	x_8
x_1	—		
x_2	(ev)		
x_3	(dVeV)		

$$f = \{ \wedge \{ V C_{ij} \} \mid i, j \leq n \}$$

~~reduce~~

This f^n will differentiate any x_i from any x_j .

But we want to differentiate only accept ones from reject ones.

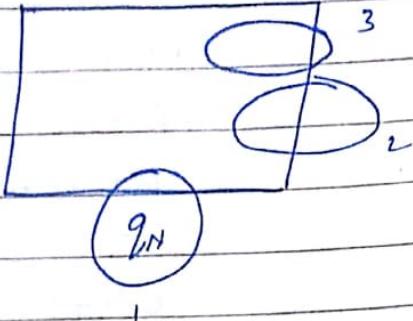
If we get only a
anywhere, it will
absorb all containing
a.

Soft Computing

→ Lower Approximation of Decision to features

Approximations of decision class

$[x]_B \rightarrow$ Equivalence Class.



We require some Degree of Overlap or Degree of membership.

How much are feature classes approximating decision class
Aim → Increase Accuracy. (Make more inside & outside)

POSITIVE Set \rightarrow Union of all lower approximation.

$\frac{|\text{Positive Region}|}{|\text{Universe}|}$

$$\frac{|\text{POS}_B|}{|\text{U}|} = \text{Dependency on feature set B.}$$

\therefore We assume 1 outside & 3 inside to make Positive Region larger.

↳ Approximation

This ratio shows Dependency of system on B.

Information.

Degree of Membership

$$\mu_X^B(x) = \frac{|[x]_B \cap X|}{|[x]_B|}$$

for 1 → 1/10 membership

for 3 → 9/10 membership.

	Age	Upbringing	Like FB?
1	0-10	Rural	Y
2	10-20	Urban	Y
3	10-20	Semi	Y
4	10-20	Rural	N
5	20 onwards	Urban	N

$$IND_{Age} = \{ \{1\}, \{2, 3, 4\}, \{5\} \}$$

$$\mu_Y^{Age}(1) = 1$$

$$\mu_N^{Age}(1) = 0$$

$$\mu(2) = \mu(3) = \mu(4) \quad \begin{matrix} \nearrow 2/3 Y \\ \searrow 1/3 N \end{matrix}$$

$$\mu_Y^{Age}(5) = 0$$

$$\mu_N^{Age}(5) = 1$$

inside or overlaps

earlier, if ~~x~~, it is a member

Now, degree of membership threshold

$$B(X) = \{x \mid \mu_x^B(x) \geq \pi\}$$

$$\bar{B}(X) = \{x \mid \mu_x^B(x) > 1 - \pi\}$$

Earlier π was 1, now will be lower.

Rough Dependency, γ
↓
 $\gamma_R(\text{Age}, D)$

$D \rightarrow$ Set of Decisions.

Rough
Dependency of Decision 0

w.r.t Set of Features (Age)

$$\text{If } \pi=1 \rightarrow \gamma = 2/5 = 40\% \quad \text{w.r.t Age}$$

If w.r.t Both, $\gamma = 1 \Rightarrow$ Dependence on less features.

$$\text{If } \pi = 66\% \rightarrow \gamma = 1$$

Now, γ from 40% to 100% with reduced features
although accuracy & hence false positives & negatives.
but without losing much info, reduced complexity.

Now Rough Reducts \therefore Dependency more imp.

Also as we use Meta Heuristics \therefore not global optimum.

Significance of feature

Taking all features \rightarrow dependence 1

We remove a feature, if dep \downarrow more \Rightarrow more

significant

$$\sigma_{C,D}^{(a)} = \frac{\gamma(C, D) - \gamma(C - \{a\}, D)}{\gamma(C, D)}$$

$$\sigma_{C,D}^{(B)} = \frac{\gamma(C, D) - \gamma(C - B, D)}{\gamma(C, D)}$$

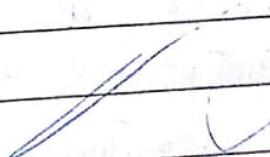
We can upto a limit do rough Reduct, when manageable apply analytical method.

Error in approximate Reducts

$$E_{C,D}(B) = \frac{\gamma(C, D) - \gamma(B, D)}{\gamma(C, D)}$$

B is an approximate Reduct if this error is acceptable
i.e. errors < threshold (say 5-10%)

If error ↑, either replace a feature with another or add new features.



If domain real.

SC

How to discretize data? (Like in age 10-20 20-30)

↳ will lose information

We must discretize w.r.t attribute decision

We need decision based clustering.

without decision
↓
lose information.

(can use K-means clustering.)

We put seed data.
Eg → marks.

Seeds → 8, 16, 20
Randomise

Threshold → $\pm 7 \rightarrow$ will auto

that seed

Centroid keeps changing

Each object → Bedsheet

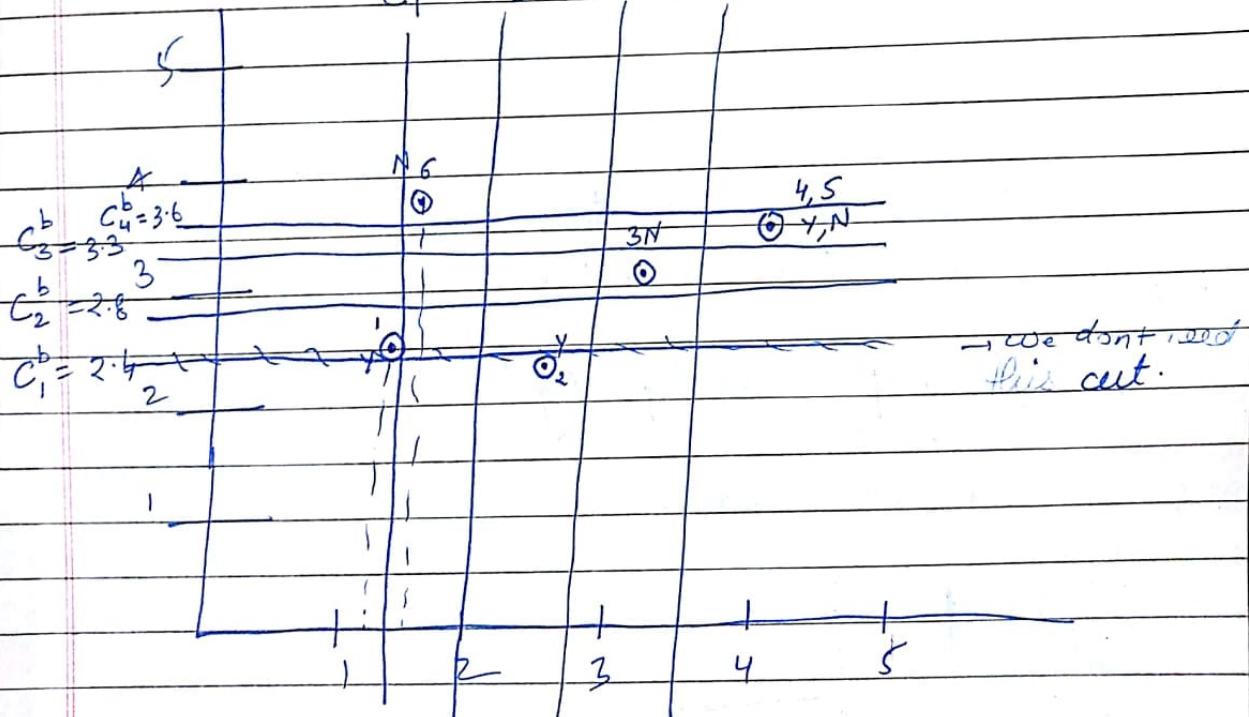
Table on grp.

Decision → covers Bed or not?

dist. Based Space : length vs Breadth.

we sort value

$$C_1^l \quad C_2^l = 2 \quad C_3^l = 2.85 \quad C_4^l = 3.65$$



We want to partition this space.

Cut b/w closest pts.

$$\frac{1+2+1.5}{2}$$

→ cut along length direction.

we need min. cuts to discern.

We can use analytical method as done for reducts

but not used as we don't require all

ensemble learning → We want to keep all reducts

get 1 set of reducts here, we can take only one set of cuts (min.)

(if features missing)
(enrich rule base)

(MD) Maximum Discernibility Algorithm

(Col) $c_1^l c_2^l c_3^l c_4^l c_1^b c_2^b c_3^b c_4^b$

$U_1 U_3$ 1 1 1 1

If a row has only 1, take that cut surely
Otherwise take max-cut ~~all~~ columns.

$f = C$

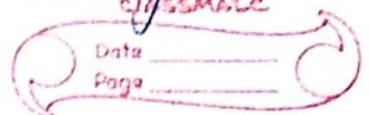
$\underline{(f_3^b \wedge f_3^l) \text{ or } (f_3^b \wedge f_2^l)}$

Taking this table. →

f_3^b	↑	1	
	↓	0	1
		1	

Literature Survey.

classmate



↳ length breadth d

1	0	0	Y
2	0	0	Y
3	1	0	N
4	1	1	Y
5	1	1	N
6	0	1	N

codified the
space

(Discretization)
for real values

we are removing infoⁿ which does not contribute to classification.

Decision Relative Discretizing.

We make 2 graphs for a & b & connect those which need to be separated.

This edge signifies $b_1 \& b_4$ need to be given diff. codes.

SC

Training Data \rightarrow Table

Training \rightarrow Making rules from Table (Database
once we have Information System,

- 1) Ensure features needed are there in database (have all condition attributes needed) (can have more than reqd)
- 2) Cleaning data & filling missing data

How to fix? \rightarrow Fill with most frequent data of column.

Make equivalence classes without that feature
& fill with most frequent of its equivalent class
can see decision & fill with similar decision.

- 3) Discetize
or
(if too much Data)

Clusters \rightarrow not good. \rightarrow not decision related
 \rightarrow centroid changes

\rightarrow MD Algorithm.

\rightarrow features replaced by codes.

- 4) finding reducts

Rough Membership Threshold

Remove features with low error of approx.

Find Rough reducts

when reduct manageable \rightarrow Boolean logic (Discernibility matrix)

- \Rightarrow Make equivalence classes &

In general, why divide parallel to axes?
Can take a hyperplane

$$H = \alpha_1 a_1 + \alpha_2 a_2 + \dots + \alpha_i a_i + \dots + \alpha_{k+1}$$

We need to find α 's to find Hyperplane.

AIM → To discern.

Ideally → All with one decision should be on one side
of hyperplane.

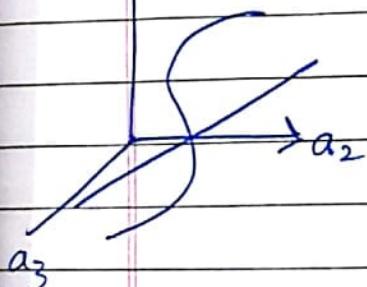
can use GA for this to calculate α 's.

$$\text{Conflict}(R) = \frac{1}{2} | \{u_i, u_j\} \mid d(u_i) \neq d(u_j) |$$

How many pairs of object
← differ in their
decision.

Info we are interested in
(differentiating classes)

q1. If powers also
of attributes



$$C_i^L(H) = \{u : \sum_K \alpha_K a_K(u) < 0 \quad u \in X\}$$

$$C_i^R(H) = \{u : \sum_K \alpha_K a_K(u) \geq 0 \quad u \in X\}$$

Find these for every decision class.

$$L = |C_i^L(H)| \quad R = |C_i^R(H)|$$

Award Fⁿ

Turn
Page

$$\rightarrow \sum_{i \neq j} |C_i^L(H)| / |C_j^R(H)|$$

AIM \rightarrow Maximize this.

A^{II} \rightarrow of same.

$$\text{Award}(H) = LR - \sum_{i=1}^n L_i R_i$$

$$\sum_{i=1}^n L_i \quad R = \sum_{i=1, n} R_i$$

\hookrightarrow Denotes energy of Hyperplane \rightarrow needs to be minimized.

Entropies \rightarrow

Sum-Minority or Max-Minority

$$\min \{|L_i|\} + \min \{|R_i|\}$$

$$\max \{\min\{|L_i|\}, \min\{|R_i|\}\}$$

\downarrow
Minimize this.

Sum-impurity

$$\sum |L_i| \times (\bar{x}_i - \text{avg}_L)^2$$

$$+ \sum |R_i| \times (\bar{x}_i - \text{avg}_R)^2$$

$|X_i| \rightarrow$ size of population of that i^{th} decision class.

~~ideally~~ $\rightarrow |X_i|$ of i^{th} decision class

$$\text{avg}_L = \frac{\sum_i |X_i| / |L_i|}{\sum_i |L_i|}$$

Penalties

$$\text{Penalty}(H) = \sum_{i=1}^k L_i R_i$$

or a mix of Award & Penalty

$$w_1 \text{award}(H) - w_2 \text{penalty}(H)$$

$$w_1 \text{award}(H)$$

$$w_2 \text{penalty}(H) + w_3$$

chromosome Encoding for GA \rightarrow Values of α 's.

we are partitioning & discretizing, but now not II to axes
 & put only 1 oblique hyperplane.

we can make feature wise hyperplanes
 or for ~~feature~~ of subset of features

less game playing as
 class separable

SVM

Quality Metrics for Rules

a) Support

Support is same as Dependence.

$$\text{Dependence } \gamma(C, D) = \frac{|\text{Pos}(C, D)|}{|U|}$$

of a decision
on set of features

To measure support of a rule, (basically an object is a rule)

$$f_x(C, D) = \frac{|[x]_c \cap D(x)|}{|U|}$$

objects which have same decision & same equivalence class

Partial Dependence of D on C w.r.t a given object equivalence class.

Certainty of a Rule

$$Cer_x(C, D) = \frac{|[x]_c \cap D(x)|}{|[x]_c|}$$

$$Cer_x(C, D) = \frac{Y_x(C, D)}{|[x]_c| / |U|}$$

$$Cer_x(C, D) = \frac{Y_c(C, D)}{\text{TF}([x]_c)}$$

→ Proportion of popⁿ for this equivalence class.

$$Cer_x(C, D) = \frac{Y_c(C, D) / U}{|[x]_c|} \quad (\text{Ratio Proportionate})$$

If $[x]_c$ was entire popⁿ, what would be dependency?
Certainty

Coverage of a rule.

Pawlak

Pattern of RST
in paper

$$Cov_x(C, D) = \frac{|[x]_C \cap D(x)|}{|D(x)|}$$

$$Cov_x(C, D) = \frac{\# f_x(C, D)(U)}{|D(x)|}$$

		How much of decision is covered		Usage
Segment	change in services	incoming calls	outgoing calls	
Seg	Inc	Out	change	Churn
1	Med	Med	Low	No
2				200
3				
4				
5				
6				

→ No of People
with these
features

Grop

Given some statistics, we must be able to derive other statistics.

→ Predict
~~Quality of classified~~
Recall confusion ~~Metrics~~ Metrics.

Precision

Accuracy.

✓ F1 Measure