Date

Page (1)

UNIT 1

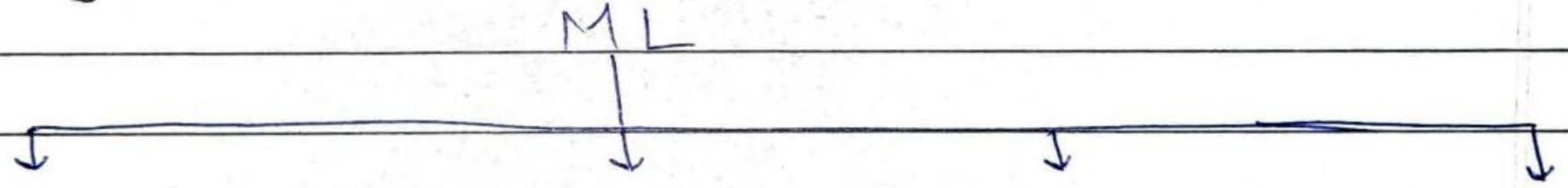
Machine Learning

Introduction:-

→ Machine learning. UNIQUE APPSITES

* application of AI, ability to learn and improve from experience
* Learn without human assistence.
* Analysis of massive quantity of data.
* Fast & accurate results.
*

→ Types:-



(1) Supervised learning: trained on labelled dataset, have both I/p & O/p parameter.

(2) Unsupervised: Not trained on labels, uses ML algo to analyze cluster unlabeled datasets, No right O/p.

(3) Semi: Improve learning accuracy. Trial, error search, delayed reward, agents automatically determine ideal behaviours & max performance.

(4) Reinforcement: Produce action, discover error, Trial, error

→ Adv of Machine learning algo:-

(1) Supervised.

* classes represent features * Training data is reusable

Disadvantages Of supervised :-

* Classes may not match spectral classes.
* Consistency vary * Training data need cost & Time.

(2) UnSupervised :-

UNIQUE APPSITES

Adv:- * No prev knowlege req * Easy & fast
* Uniques spectral classes * human error is minimized

Dis: * Do not present features * No spatial relationship
* Take time to interpret spectral classes.

(3) Semi-supervised :- Adv * Easy to understand * Simple.
* Stable & fast * high efficiency * less annotated data needs.

Dis: * Iteration result is not stable * Not for network level data
* low accuracy.

(4) Reinforcement : Adv: * Solve complex problems
* achieve long-term results * Similar to learning of human.
* close to achieve perfection.

Dis: * overload states & diminish result
* Not for solving simple problem
* Need lot of data & computation.
* Curse of dimensionality.

→ Applications: Image recognition, speech recog, Medical diagnosis, extraction, etc.

→ Adv of ML:- * Patterns & trends identification.
* Automated * Continous improvement
* Handle Multi-dimentional & variety of data.

→ Dis of ML: * Data acquisition * Time & resources
* Interpretion of result * High error susceptibility

→ Well defined learning prob:-

* 3 traits → Task (T)
→ Performance Measure (P)
→ Experience (E).

* A program said to learn from E in context to some T & some P, if its performance at tasks in T, as measured by P, improved exp^n E.

Ex. A checkers learning problem.
(a) T = Playing checkers
(b) P = % of game won.
(c) E = Playing practise games against itself.

Role in ML:- UNIQUE APPSITES

(1) Learning to recognize spoken words:-

* Speech recog: * SPHINX system.
* NNL & Makov model is effective for auto customizing speakers, vocabulary, mic, bg noise, etc.

LIKE, SHARE & SUBSCRIBE

(2) Learning to drive an autonomous vehicle:-
* ALVINN system (70 miles/hr for 90 miles). * Sensor based

(3) " " Classify new astronomical structures:-
* large databases * decision tree learning algo
* Auto classify all objects * Has 3TB of image data.

(4) " " play world class backgammon:-
* TD-GAMMON (played 1M+ practice games against itself).
* Competitive with human WC * large search spaces examined easily.

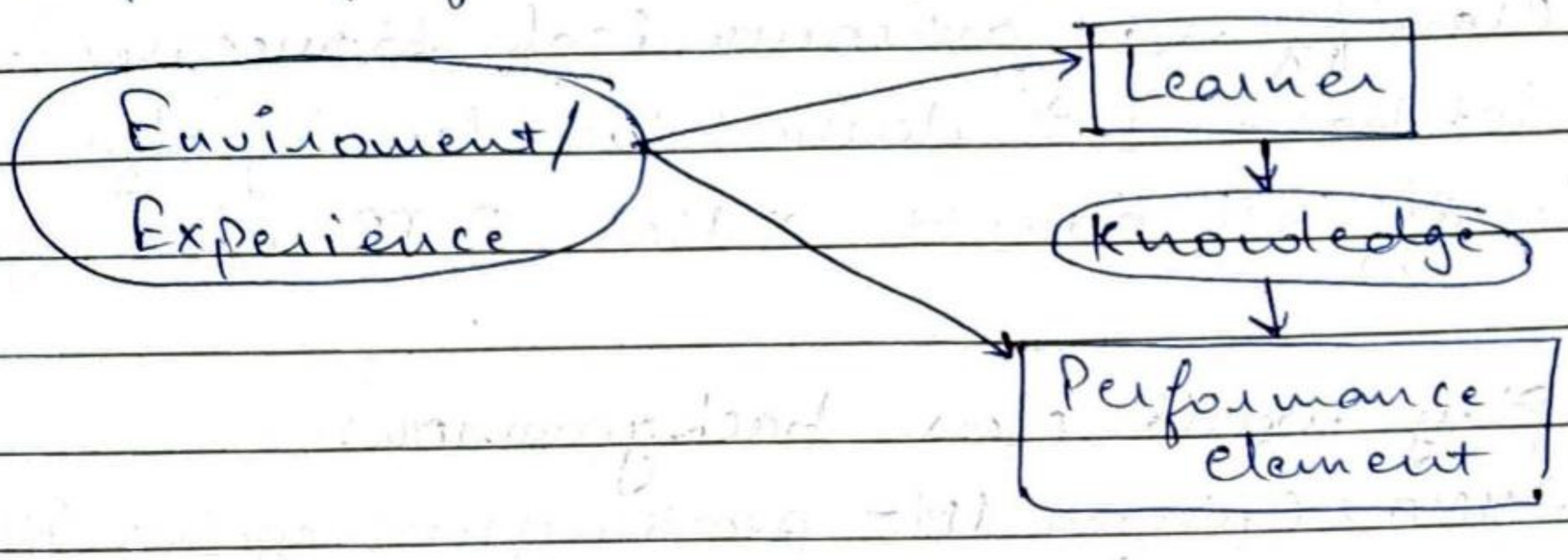
→ AI vs ML:- UNIQUE APPSITES

	AI	ML
①	Human intelligence, perform simple to complex task.	① Provide ability to learn & understand.
②	Do task in human way or smart way	② Teach Computers to think & understand like humans.
③	Based on human intelli	③ Based on system of probability.
④	Healthcare, finance, media, etc.	④ Web sec, recog, learning, etc.

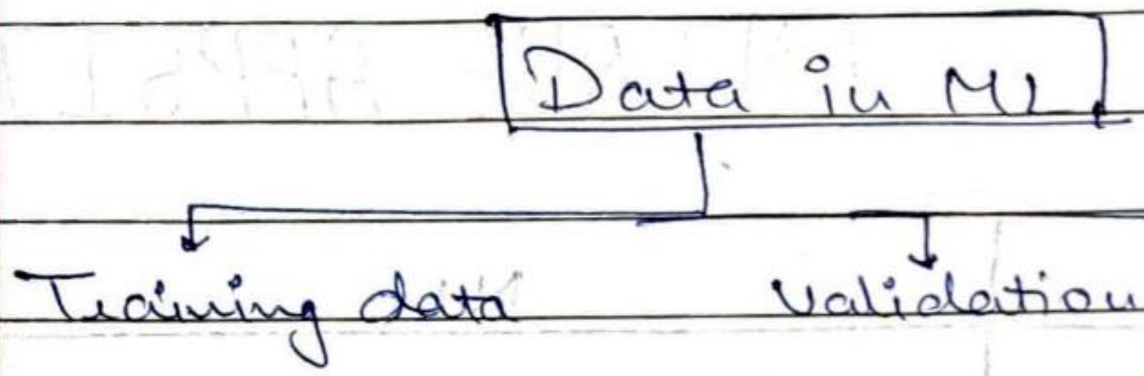
→ Steps to design learning system:-
* Specify learning task.
* Choose set of training as per training exp.
* Divide training data in groups/classes & labels.
* Determine type of knowledge representation to be learned from training exp.
* Choose classifier to generate general hypothesis

from training data.

* Apply learner classifier to test data.
* Compare performance of system with expert human



→ Data Splitting :-

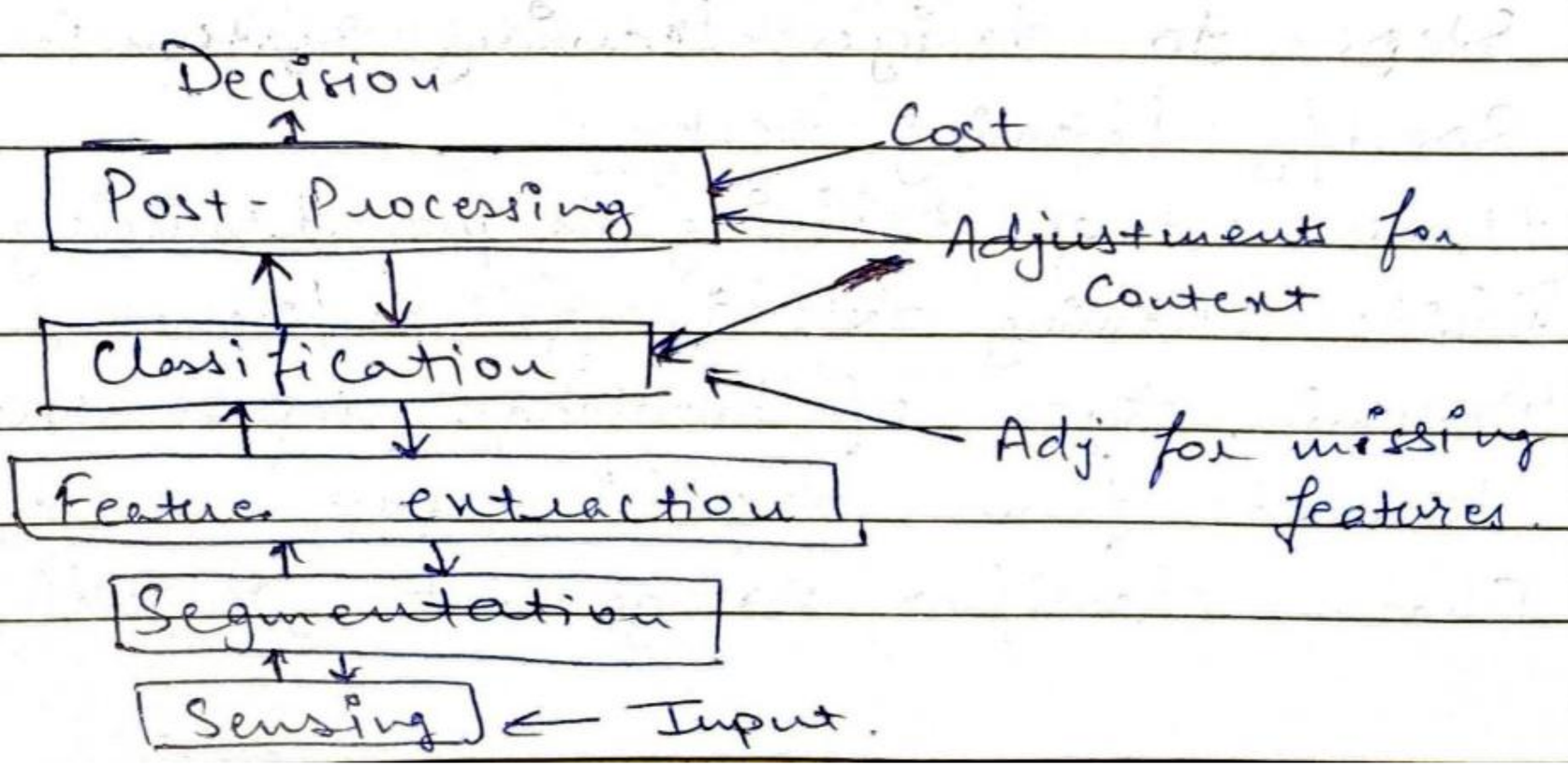


→ Terminologies :-

* Features * Feature vector * Feature Space
* Class (w) * Decision boundary * Classifier
* Error * Training performance * Generalization.

→ Components :- UNIQUE APPSITES.

* Sensing * Segmentation * Classification
* Post processing * * Feature extraction.



FREEMIND

Date

Page (6)

(r

→ Find S- algo:-

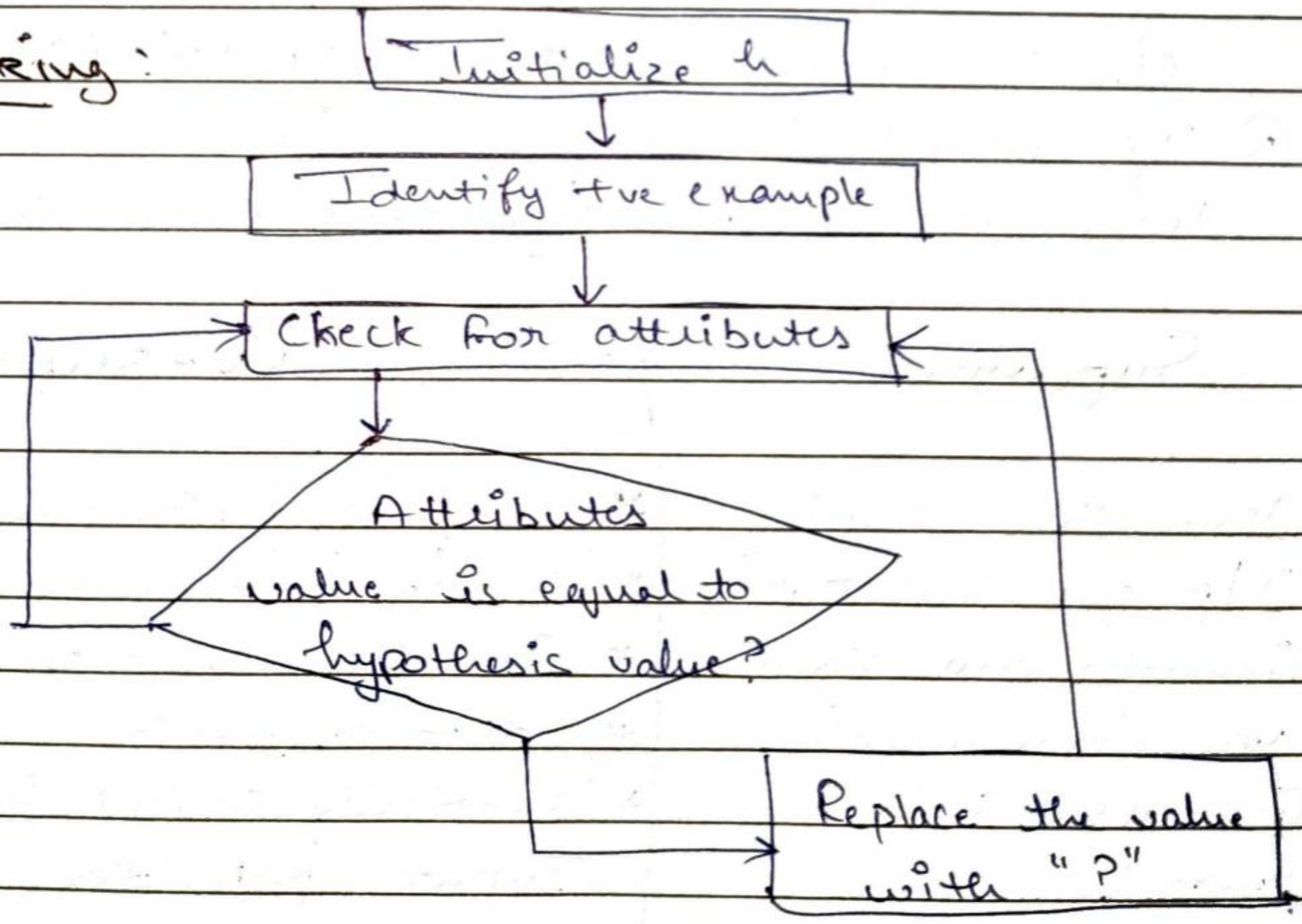
Algo:- (1) Initialize h to most specific hypothesis in H.
(2) for each +ve training instance x.
For each attribute constraint a_i in h.
IF Contraint a_i in h is satisfied THEN
DO NOTHING
ELSE
replace a_i in h by next more general constraint satisfied by x.
(3) O/p hypothesis h.

UNIQUE APPSITES

Adv: If training data is correct, then it can guarentee O/p most specific in hypothesis H.

Dis: * Can't determine hypothesis is consistent through data.
* data can mislead algo * No back tracking tech support
* learning process is poor * Robustness to noice is weak.

Working:



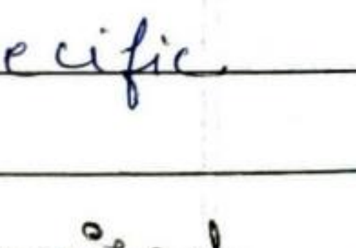
→ Candidate elimination algo:-

* Extended form of find-s algo.
* Consider +ve & -ve examples.
* +ve example considered as find-s algo, where -ve " . Specified from generalize form.

* It incrementally builds version space given a hypothesis space H & set E of examples. The ex are added one by one, each ex possibly shrinks the version space by removing hypothesis that are inconsistent with ex.

UNIQUE APPSITES

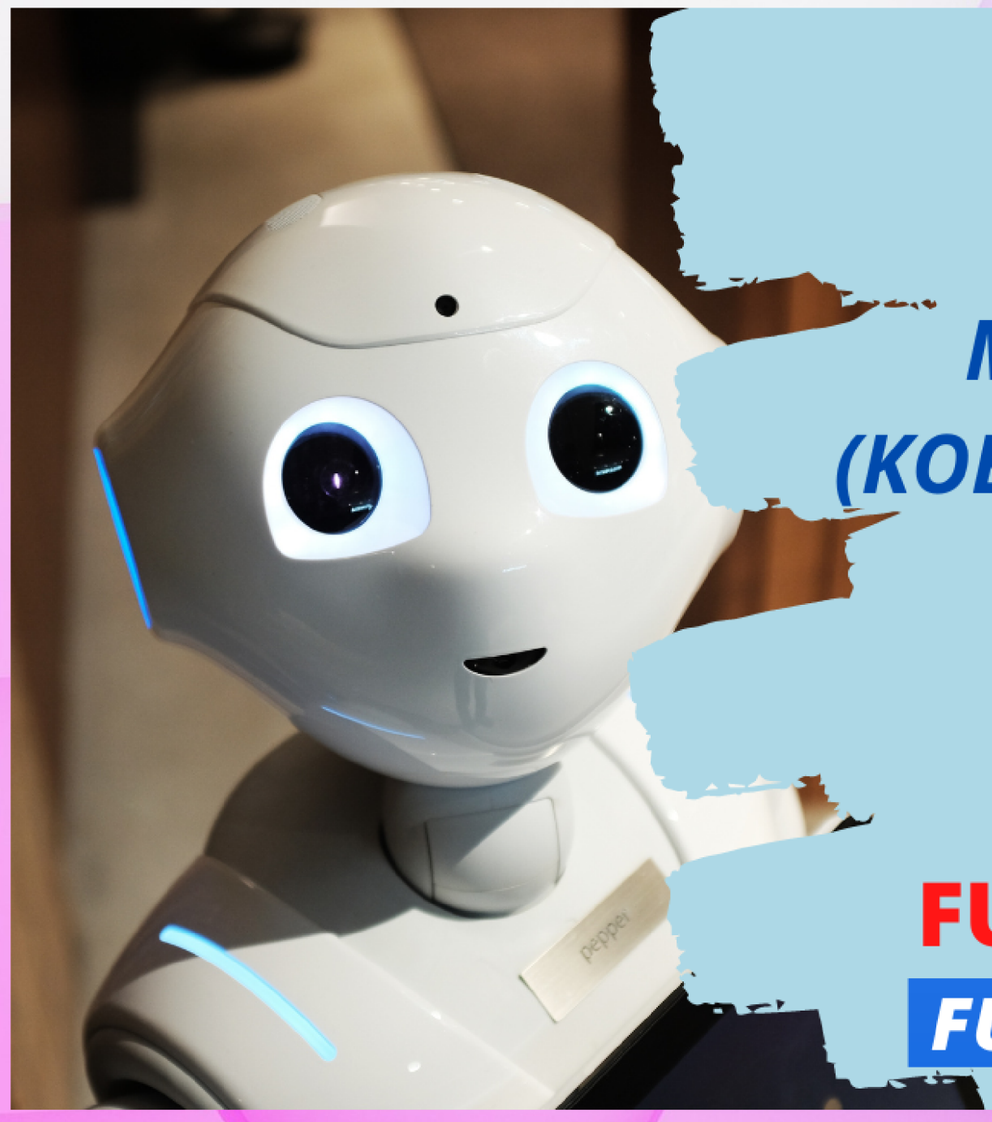
Algo:- ① Load data set
② Init Gen hypothesis & specific hyp"
③ For each training example
④ If example is +ve example
if attribute_value == hypothesis_value:
Do nothing.
else: replace attribute with '?' (Generalizing it)
⑤ If example is -ve.
Make generalization hypothesis more specific.



→	Supervised	Un-supervised
*	Associative learning	* Self-organisation.
*	I/p for goal o/p	* Neural Networks.
*	Generates error signal	* Generate new class.
*	Non-linear mapping	* data compression & clustering.
*	Gen global & local model.	* statically salient features of Input population.

UAS
Unique AppSites

LIKE
& SUBSCRIBE



MACHINE LEARNING
(KOE-703/RCS-080/ROE-083)
AKTU / UPTU
UNIT 1
(INTRODUCTION)
FULL EXPLANATION

FULL EXPLANATION Unit 1

CLICK HERE

BY ANUBHAV GOEL

Date

Page 1

UNIT 2

MACHINE LEARNING

DECISION TREE LEARNING

→ Contents

* Decision tree.
* Decision tree learning algorithm
* Artificial Neural Network (ANN)
* Types of Neuron Connection with architecture.

→ Decision tree :-

* It is a flowchart like representation of data that graphically resembles a tree that has been drawn upside down. In this, the root is a decision that has to be made, the tree's branches are actions that can be taken & the tree's leaves are potential decision outcomes.

* The paths from root to leaf represent classification rules.

* It is a predictive modelling approach used in statistics, data mining & machine learning.

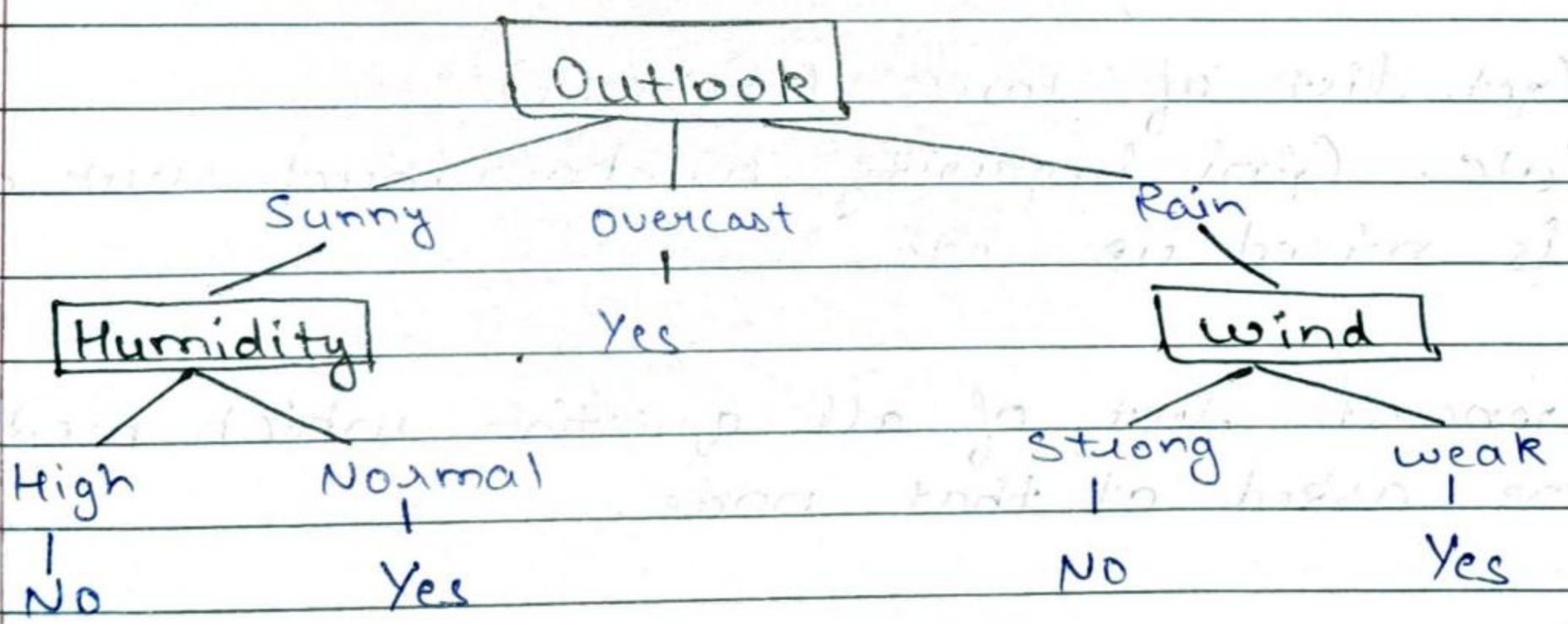
* Constructed via an algorithmic approach that identifies the ways to split a data set based on diff conditions.

* Non-parametric survised learning method used for both classification & regression tasks.

* Classification trees are tree models where the target variable can take a discrete set of values

* Regression trees are decision trees where target variable can take continous set of values.

BE * Decision making with labels (Rain(Yes), Rain(No)).



→ Advantages :-

UNIQUE APPSITES

* Generate understandable rules.
* Perform classification without requiring computation.

LIKE, SHARE & SUBSCRIBE

* Handle both continous & categorical variable
* Provide clear indication for the fields that are important for prediction or classification.

→ Disadvantages:

* Less appropriate for estimation tasks where goal is to predict the value of Continous attribute.

* Prone to errors in classification problems with many class & relatively small no. of training ex

* Computationally expensive to train.

→ Steps used for making decision tree.

(1) Get list of rows (dataset)
(2) Calc. Gini impurity or how much our data is mixed up, etc.

(3) Generate list of all question which need to be asked at that node.

(4) Partition rows in True or false based on the question asked.

(5) Calc. info gain based on Gini impurity & partition of data from previous step.

(6) Update highest info gain based on each ques.

(7) Update ques based on info gain.

(8) Divide the node on ques. Repeat again from step 1 until we get pure node. ~~If~~ (leaf node)

→ Gini impurity and index: UNIQUE APPSITES

* It is a method focuses on purity and impurity in a node

* Gini impurity index measures the impurity of an input feature with respect to the classes.

* Gini impurity index reaches its minimum (zero) when all attributes in the node fall into the classes.

* Tree spliting is based on choosing the attribute with lowest index of split.

* $Gini(t, X) = \left(\frac{n_1}{N_t}\right) I_G(t_{x(x1)}) + \dots \left(\frac{n_r}{N_t}\right) I_G(t_{x(x_r)})$

→ Issues related to application of decision tree.

(1) Missing data: (a) when values have gone unrecorded or they might be too expensive to obtain.

(b) * To classify an object that is missing from test attributes.

* To modify info gain formula when example have unknown values for attribute.

(2) Multi-Valued attribute:

* when an attribute has many possible values, the info gain measures gives an inappro-priate indication of attribute's usefulness.

* we could use an attribute that has a diff value for every example.

(3) Continuous and integer valued input att":

(a) Height & weight have an infinite set of possible values.

UNIQUE APPSITES

(b) Split the point that gives highest info gain.

(c) Efficient dynamic programming method exist to find good split point but it is the most expensive part of real world example.

(4) Continous values output attributes:

(a) To predict numeric value we need regressiontree.

(b) Such tree has linear func" of some subset of numerical att", rather than a single value at each leaf.

(c) Algo must decide when to start & stop applying linear regression using remaining attributes.

→ Basic terminology used in decision tree.

(1) Root Node: Represent entire population & further divided in two or more homogeneous sets.

(2) Splitting: Divide one node in sub-nodes.

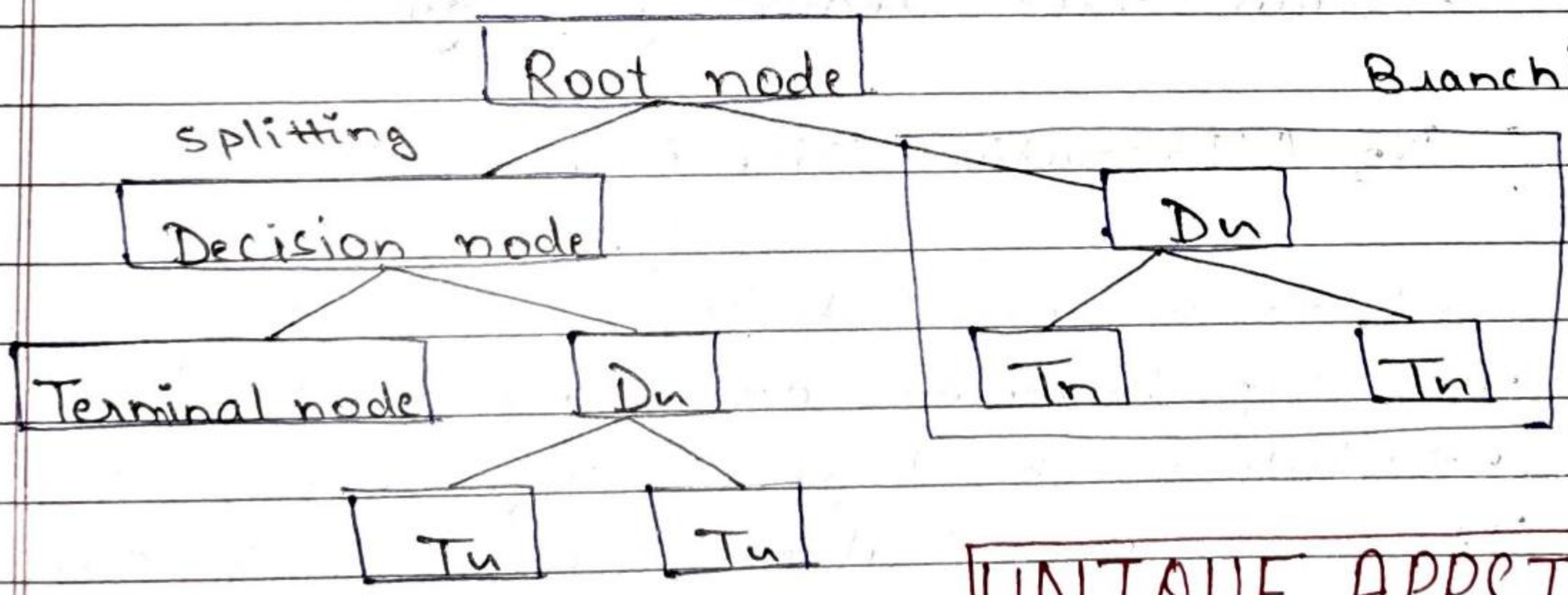
(3) Decision node: Dividing sub-nodes to sub-nodes.

(4) Leaf / Terminal node: Nodes that do not split.

(5) Pruning: Removing of sub-nodes.

(6) Branch / Sub tree: Sub section of entire tree.

(7) Parent & child node: Node which divide into subnod is parent and sub-nodes are child node.



→ Why do we use decision tree.

* Visualized, simple to understand, interpret.
* Require less data prepration.

* Handle both categorical & numerical data.

* Handle multi-output problems
* Explain by boolean logic (Yes or No)
* Used even if assumptions are violated by dataset from which data is taken.

UNIQUE APPSITES

→ Application of decision tree in various areas of data mining.

* E-commerce : generate online catalog
* Industry : producing quality control.
* Intelligent vehicles.
* Medicine : diagnose disease
* Business : used in CRM and etc.

→ Decision tree learning algos:

(1) ID_3 (Iterative Dichotomiser 3):

(a) Used to generate decision tree (Dt) from dataset
(b) Uses top to down and greedy search.
(c) Attribute with highest info gain is selected
(d) It is heuristic algo because it can't construct the smallest tree

Adv: (a) training data is used to create understandable prediction rules
(b) Builds short & fast tree
(c) Searches whole dataset to create whole tree.

(d) Find leaf nodes and enable to prune, to reduce no. of tests
(e) The calc. time is linear func" of product of characteristic no. & node no.

Dis: (a) For small sample, data may be overfitted or overclassified.

(b) Consumes a lot of time to make a decision.
(c) large no. of input may affect features.

(2) C4.5:

(a) An algo used to generate DT, an extension of ID3
(b) It is a statical classifier.
(c) Better than ID3 as it deals with continous and discrete attributes and also with the missing values & pruning trees after construction
(d) C 5.0 is commercial successor of C4.5 because it is faster, m/m efficient & used for building smaller decision tree.
(e) C4.5 performs tree pruning process by default.

Adv:- (a) Easy to implement
(b) C4.5 build models which can be easily interpreted.
(c) Handle both categorical and continous values.
(d) Deal with noise and missing value attributes.

Date

Page 9

Dis of C4.5 :-

(a) Smaller variation in data can lead to diff DT
(b) Not work well for small training set.

(3) Classification & Regression Tree (CART):

(a) Constructed by binary splitting of attribute.
(b) Gini Index is used for selecting splitting attr
(c) Used for regression analysis by regression tree.
(d) Has an avg. speed of processing & Support both continous & nominal att" data.

Adv :- (a) Can handle missing values automatically using proxy splits.
(b) Uses combination of continous / discrete variables
(c) Automatically performs variable selection.
(d) Establish interactions among variables.
(e) Doesn't vary acc. to monotonic transformation of predictive variable.

Dis :- (a) Unstable ~~for~~ trees
(b) Splits only one variable
(c) Non-parametric algo.

→ Artificial Neural Network (ANN):-

* They are computational algo that intended to simulate the behaviour of biological system composed of neurons.

* Inspired by animal's central nervous system.
* Capable of ML as well as pattern recognition.

* Neural Network is an oriented graph. It consist of nodes which in the biological analogy represent neurons, connected by arcs

* Corresponds to dendrites & Synapses.
* NN is a ML algo based on human neuron.
* Sends & process signals in form of electrical and chemical signals.

* Neurons are connected with Synapses.
* ANN is an info processing technique.
* ANN includes a large no. of connected processing units that work together to process info and generate meaningful results from it.

* A neural network contain 3 layers.

(a) Input layer: Raw info that can feed in network.

(b) Hidden layer:
(i) determine activity of each hidden unit.
(ii) Activity of I/p units & weights depend upon connection b/w the I/p and hidden units.
(iii) One or more hidden layers possible.

(c) O/p layer: Depends on activity of hidden units & weights b/w the hidden & O/p units.

Date

Page 11

→ Adv of ANN :-

(a) Problems represented by attribute - value pairs

(b) used for problems having target func", O/p may be discrete-valued, real valued, or a vector of several real or discreate valued att'.

(c) Learning methods are quite robust to noise in training data, training ex may have error but final o/p is not affected.

(d) Used for fast evaluation of learned target func" required.

(e) Can bear long training times depending on factors such as no. of weights in network, the no. of training ex and settings of various learning algo parameters.

→ Disadvantages:

(a) Hardware dependence:

(i) ANN require processors with parallel processing power, by their structure.

(ii) For this reason, the realization of equipment is dependent

(b) Unexplained functioning of network:

(i) Most imp problem of ANN.
(ii) In a probing Sol, it doesn't give a clue as to why and how.
(iii) This reduces trust in the network.

(c) Assurance of proper network structure:

(i) No specific rule for determining structure of ANN.
(ii) Proper network structure is achieved through experience and trial & error.

(d) The difficulty of showing the problem of network

(i) ANN can work with numerical info.
(ii) Problem translated into numerical values before being introduced to ANN.
(iii) Directly influence the performance of network
(iv) Depend on user's ability.

(e) The duration of network is unknown:

(i) The network in reduced to a value on which error are less means training has completed.

(ii) This value doesn't give us optimum results.

→ Characteristics of ANN :-

(i) Neurally implemented mathematic model.
(ii) Contains large no of interconnected processing elements called neurons to do all ops
(iii) Info stored in neurons is basically the weighted linkage of neurons.
(iii) The input signals arrive at processing elements through connections & connecting weights.

(v) Has ability to learn, recall & generalize from give data by suitable assignment & adjusting weights.

(vi) Collective behaviour of neurons describe its Computational power and no ~~signal~~ single neuron carries specific information.

→ Application areas of ANN :-

(1) Speech recoguition :-

(a) Very important in human-human interaction.
(b) Speech interfaces make human interaction easy with computers.
(c) Humans still need sophisticated languages to communicate with Computers in present era.

(d) To make communication, a simple sol" could be Commu" should be in spoken language

(e) Hence, ANN is playing a major role in speech recognition.

(2) Character Recognition:

(i) Falls under general area of pattern recognition

(ii) Many NN developed for automatic recognition of handwritten characters, either letters or digits.

(3) Signature verification application:-

(i) Useful way to authorize and authenticate

(ii) Non-vision based technique.

(iii) We have to train NN using an efficient NN algo with the feature sets like geometrical feature set representing sign.

(iv) Trained NN will classify sign as genuine / Fraud.

(4) Human face Recognition:-

(i) Biometric methods to identify the face.

(ii) Typical task as characterization of 'non-face' images

(iii) If a NN is well trained, then it can be divided in two classes namely images having face & images that do not have faces.

UAS
Unique AppSites

LIKE
& SUBSCRIBE

DR. A.P.J. ABDUL KALAM TECHNICAL UNIVERSITY, UTTAR PRADESH
LUCKNOW

MACHINE LEARNING
(RCS-080/ROE-083)
AKTU / UPTU
UNIT 2
(DEEP TREE LEARNING)
FULL EXPLANATION
FULL EXPLANATION PART 1

CLICK HERE

BY ANUBHAV GOEL

Date

UNIT 3

Machine Learning.

Evaluating Hypothesis.

→ Contents

* Hypothesis
* Bayesian learning
* EM algorithm.
* Bayesian network

→ Machine learning :- application of AI, ability to automatically learn and improve with experience, primary aim is to learn without human involvment, Analyse massive quantity of data, generate faster and accurate results, Combining ML with AI becomes effective in processing large volumes.

→ Hypothesis :- * describes target in supervised machine
* we find func with best maps input & output, cab also called func approximation.
* Tentative relationship b/w two or more variables.
* Authenticity checked of statement as true or false.

* Characterstics :-

(1) Simple & clear (2) Predictable
(3) Relevant (4) Testable
(5) Manageable.

* Importance :-

(1) Ensure accuracy and research.
(2) Helps in devising research techniques.
(3) Direction to research.
(4) Saves resources (time, money & energy)
(5) Prevents blind search.

* Types :-

Hypothesis (7)

Simple Relationship	Complex	Working	Alternative	Null	Statical	Logical
B/w two or more variable. Ex. ↑ unemployed ↑ the crime.	Among more than 2 variables in a research. Ex. ↑ poverty, ↑ Crime & ↑ illetcracy.	test & work Productive theory generated.	work on Rejected or wrong.	No rela. -tionship b/w variables Denoted by 'H'	Statically verified.	verified logically

* Difficulties → Bias in estimate (poor estimates)
→ variance in estimate (accuracy varies).

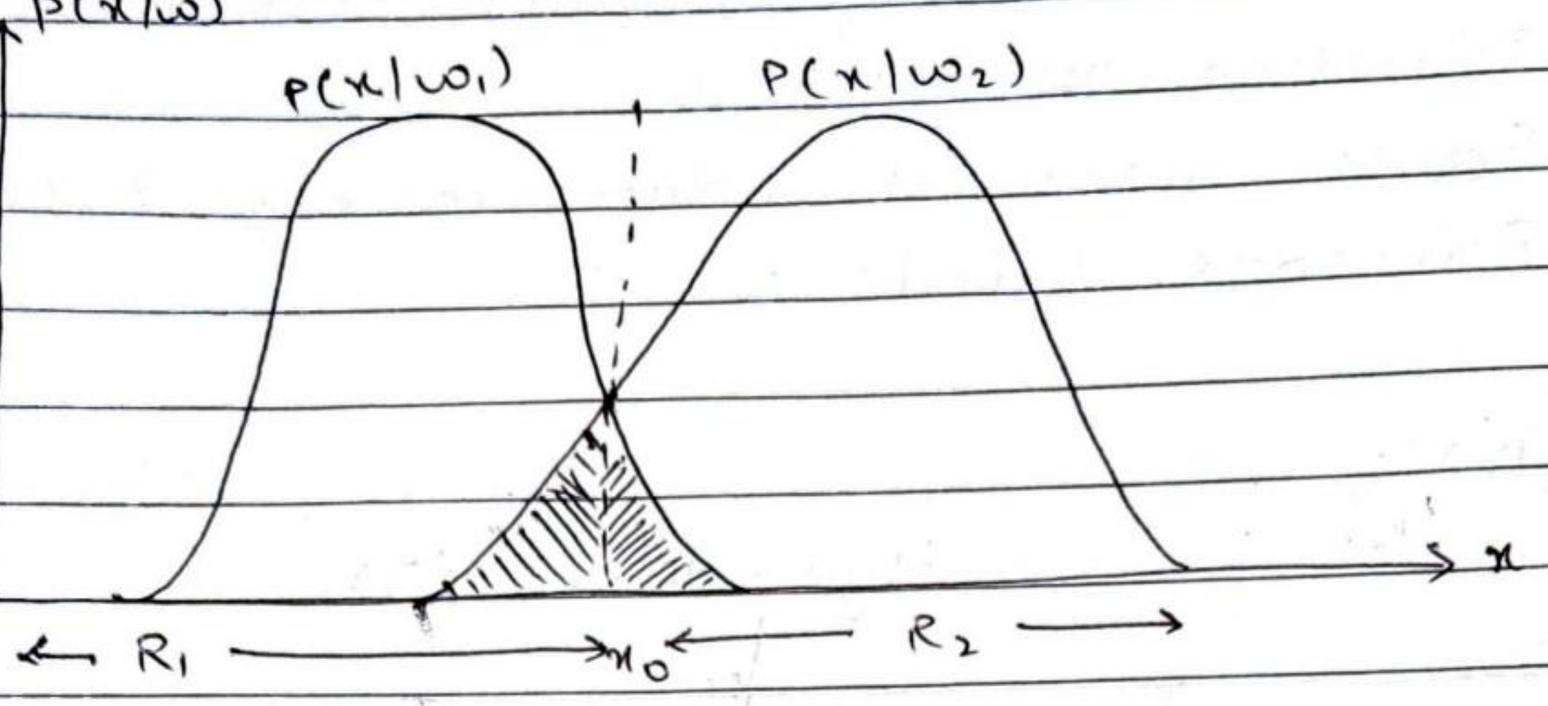
→ Diff b/w supervised & reinforcement learning: Q3.9

→ Bayesian learning :-

UNIQUE APPSITES

* Statical approach.
* Quantify b/w classification approaches.
* State nature of things in a particular pattern.
* Denote state of nature by 'ω'.
* Ex :- Consider various no. of ball of colour red and blue, when ball is red $\omega = \omega_1$ & blue $\omega = \omega_2$, as state of nature is unpredictable. Now on previous probabilities & present state of nature, the decision is made.

* Decision rule is
Decide ω_1 if $p(\omega_1) > p(\omega_2)$, else ω_2.



→ Decision error can be minimized for Bayesian classification.

* By minimizing classification error probability
* It's observed that when threshold is moving away from x_0, the shaded area increases.

* We have to decrease this shaded area to min error
* Use Bayes' rule

$$P\int_{R2} p(\omega_1|x)\,p(x)\,dx + \int_{R_1} p(\omega_2|x)\,p(x)\,dx.$$

error will be min if feature space of partioning region R_1 & R_2 are chosen.

→ Baye's Theorem :-

* Let x be a thing in pattern, where x is an "evidence". Let H be some hypothesis such that x belongs to specified class c.

* We can say, the probability that x belongs to class C is determined, given that description of x is known.

* P(H/x) is posterior probability of H on x.
* Bayes theorem is $p(H/x) = \frac{p(x/H)\,p(H)}{p(x)}$

→ Bayesian network :-

* It is a directed acyclic graph in which each node is annoated with quantitive probability information.

Specifications :-

** A set of random variables makes up the nodes of network variable may be discreate or continuous

** A set of directed links or arrows connects pairs of nodes. If arrow from x to node y, x is said to be parent of y.

** Graph has no directed cycles (directed acyclic graph)

* Bayesian network provide complete description of domain.

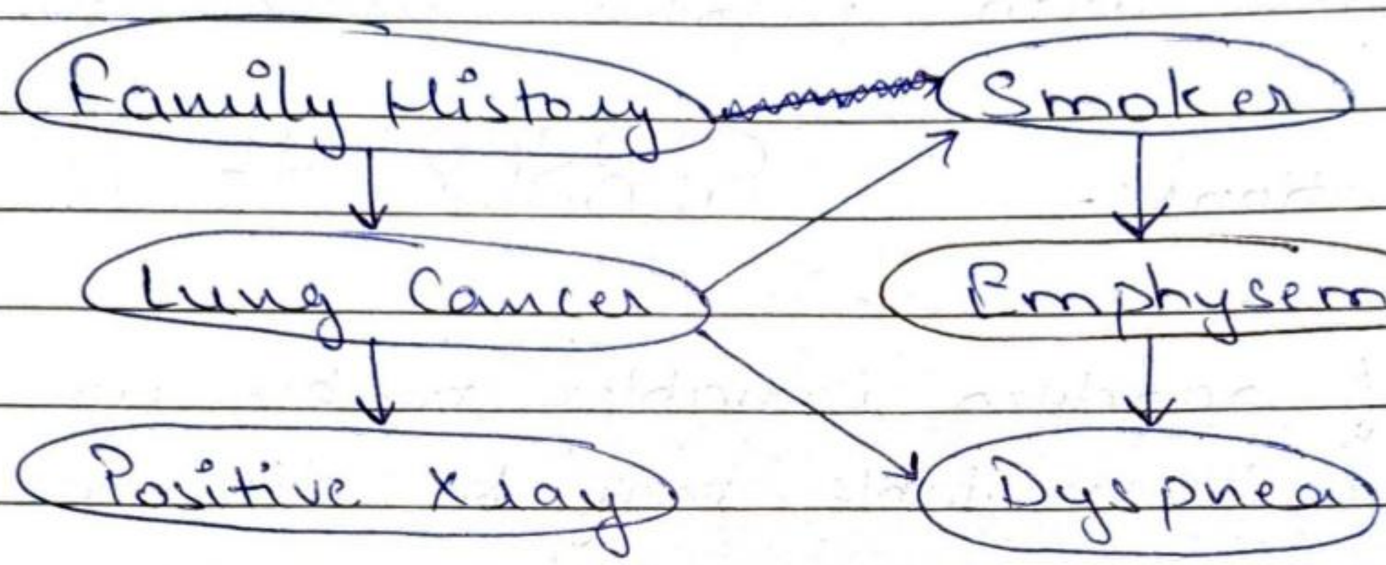
* Provide concise way to represent conditional independence relationships in domain.
* Hybrid BN includes discreate & continous variables.

→ Bayesian belief networks.

* Specify joint conditional probability distribution.
* Belief networks, Bayesian net" & probabilistic net".
* Provide graphical model of casual relationship.
* Two components:.

(a) Directed Acyclic graph

** Each node represent random variable.
** ~~Directed~~ o Discreate or continous values.
** Representation :- Six Boolean variables.



→ EM algorithm: Expectation - Maximization.

* Find max - likelihood estimates for model parameters when data is incomplete or has missing data points or has some hidden variables

* Random value choosen for missing data points and estimates new set of data.
* Find values recursively until value fixed.
* Two basic steps

UNIQUE APPSITES

(1) Estimation Step :- Acc. to given parameters, latent variables are estimated.

(2) Maximization step :- (μ_k, Σ_k, π_k)

(i) Init mean μ_k, covariance matrix Σ_k and mixing coefficient π_k, by random value.

(ii) Compute π_k values for all k.

(iii) Again estimate all parameters using presen π_k value

(iv) Compute log-likelihood func^n

(v) If log value converges to some value then stop else return to step 2.

→ Usage :-

* Fill missing data
* Used as unsupervised learning of clusters.
* Estimate parameter of Hidden Markov Model (HMM).
* Discover values of latent variables.

Adv :- * likelihood will increase with each iteration.
* E & M step is easy * M-step sol^n exist in closed form

Dis :- * Slow Convergence * local optima only.
* Requires probabilities, & forward & backward optimization.

→ Q. 3.16, 18,

Unique AppSites

LIKE

& SUBSCRIBE

MACHINE LEARNING
(KOE-703/RCS-080/ROE-083)
AKTU / UPTU
UNIT 3
(EVALUATING HYPOTHESIS)
FULL EXPLANATION

FULL EXPLANATION Unit 3

CLICK HERE

BY ANUBHAV GOEL

Date

UNIT 4

MACHINE LEARNING

Computational Learning Theory

→ Contents :-

* CLT
* Instance Based learning.
* KNN algorithm.
* Radial Basis func".
* Case based learning.
* Inductive learning problem.

→ Computational learning Theory :-

* Field of AI, determines what sort of problem are learnable.
* Ultimate goal is to understand ideas of deep learning programs, they can work or not.
* Merge : probab" theory, statics, programming opti", "info" theory, claculus & geomentry.
* It is used to : (a) Provide theoritical learning analysis.
(b) Algo expected to succeed or not
(c) when learning can be impossible.

* Categories :-

CLT

- Sample Complexity: Ex needed to find in a good hypothesis. Depends on:- Size, accuracy, Probab, Manner.
- Computational: Computational power needed to find.....
- Mistake: mistake we made to find.....

UNIQUE APPSITES

→ Instance - Based learning :- (IBL)

* Extension of KNN classification algo.
* Doesn't maintain set of abstraction of model.
* Large space req.
* Can work with noisy instance but KNN doesn't.
* Based on memorization of dataset.
* No. of parameters are unbounded & grows with size of data.
* Cost of learning process is 0.
* Also called lazy learning.

Representation : (4)

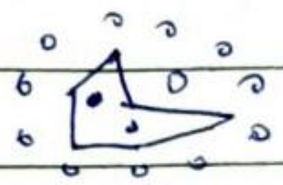
(1) IBR (1) :- * plain memorization.
* Work directly from ex rather than creating rules.

(2) IBR (2) :- * lazy * Sometime more than one NN is used & majority of class of closest KNN is assigned to new instance, this is known as KNN method.

LIKE, SHARE & SUBSCRIBE

(3) IBR(3):- * Compute distance b/w two examples.
* Standard Euclidean dis" may be used.
* identical value then distance is 0, else 1.
* Not imp to store all training instances.

(4) IBR(4):- * Don't make explicit struct" from learning



* Space are more stable with regard of class than others.

* Performance dimension of IBL

(1) Generality (Boundary) (2) Accuracy.
(3) Learning rate (4) Cost
(5) Storage requirement.

* Functions (3)
(1) Similarity func":- Similarities are numeric-valued
(2) Classification func":- Records instances of desc" for (i).
(3) Concept desc" updater: Modify concept description.

* Adv:- * Efficient * Noise resistant.
* Rich representation * Easy to understand.
* Learning is trivial.

* Dis:- * Need lots of data * Comp cost is high.
* Restricted to $x \in R^n$ * Req large m/m.
* Implicit weights of attributes.
* Expensive application time.

→ KNN algo :

* Used to decide new instance belongs to which class.
* When K=1, we have NN algo
* Classification is incremental.
* Training phase doesn't exist as all instances are stored, training used to find neighbours quickly.
* During testing, KNN class" algo has to find KNN of a new instance, time consuming if exhaustive comparison is done
* Local neighbour used to obtain prediction.

Algo: Let No. of training data sample = ~~m~~ n
p be an unknown point.
(1) Store training data samples in array of data points array, represented by tuple (x_i, y_i).
(2) For i=0 to m, calc" Euclidean dis" d(arr[i], p)
(3) Make set S of k smallest dis" obtained. Each distance corresponds to already classified data point
(4) Return majority label among S.

UNIQUE APPSITES

* Adv:- * No training period (lazy learner)
* Accurate
* Easy to implement

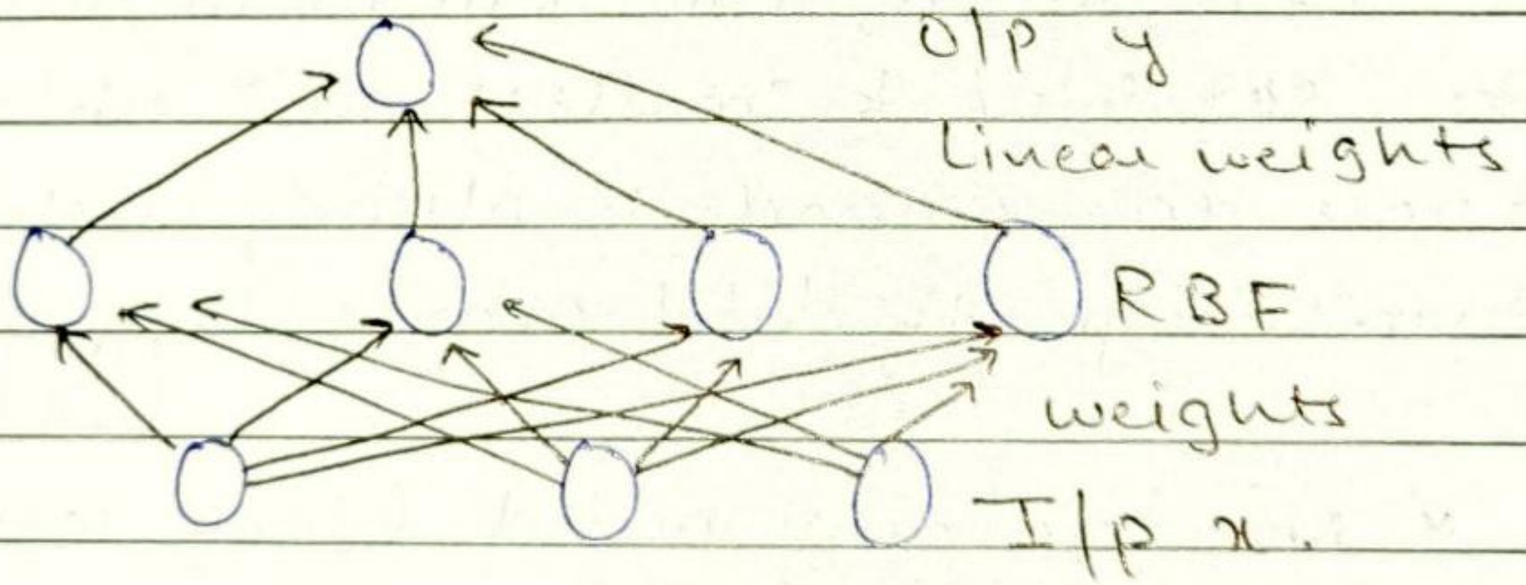
* Dis:- * Doesn't work with large data set.
* " " " " " high dimensions.
* Need feature scaling. (standardization & normalization)
* Sensitive to noisy data, missing values and outliers.

→ Radial Basis function (RBF).

* Assign real value to each input from its domain & value produced by RBF is always absolute, its measure of distance & can't be -ve.

* Euclidean distance is used.
* Used to approximate func^{n}.
* Sum represented:- $y(x) = \sum_{i=1}^{N} \omega_i \phi(\|x - x_i\|)$
* Acts as activation func^{n}.
* Architecture:- Has 3 layers.
  (1) I/p layer (2) Hidden layer (3) linear o/p layer

** The I/p can be vector of real No:- $x \in R^n$
** The o/p is scalar func^{n} of I/p vector.



→ Case-based learning algo:- UNIQUE APPSITES

* I/p as training cases & O/p as concept desc^{n}, used to generate prediction of goal.
* Distinguish by processing behaviour.

Disadvantages:- * Computationally expensive.
* Noise intolerant & irrevelant features
* Sensitive to training cases.
* Not simple.

* Function of CBL: #
(1) Pre-processor (2) Similarity
(3) Prediction (4) M/m updating.

* Benefits: (1) Ease of knowledge elicitation.
(2) Absence of problem solving bias.
(3) Incremental learning
(4) Suitable for complex problems.
(5) Sequential problem solving.
(6) Easy explaination (7) Easy maintenance.

* Limitations: (1) Large case basis
(2) Noisy data (3) Dynamic problem domain.
(4) Fully automatic operation (no sol^n sometimes)

* Applications: (1) Interpretation (2) Classification
(3) Design (4) Planning.
(5) Advising (diagnose problem resolving).

-> Major paradigms of ML. UNIQUE APPSITES

(1) Rote learning (2) Induction.
(3) Clustering (4) Analogy.
(5) Discovery. (6) Genetic algo = us
(7) Reinforcement.

Unique AppSites

LIKE & SUBSCRIBE

MACHINE LEARNING
(KOE-703/RCS-080/ROE-083)
AKTU / UPTU

UNIT 4
(COMPUTATIONAL LEARNING THEORY)
FULL EXPLANATION

FULL EXPLANATION Unit 4

CLICK HERE

BY ANUBHAV GOEL

UNIT 5

Machine learning:-

Genetic Algorithm.

→ Contents:-

UNIQUE APPSITES

* Genetic algorithm.
* Hypothesis space search.
* Genetic programming.
* Learn-one rule algo.
* Heuristic search
* Hill ~~clim~~ climbing algo.
* Reinforcement learning.

→ Genetic Algorithm:-

* Computerized search & opti° algo based natural genetics & natural selection.
* Convert design space in genetic space. design space is set of feasible solutions.
* work with coding of variables.
* Benefits:- (1) Robust
(2) opti° over large space ~~set~~ state.
(3) Presence of noise.

* Applications:- (1) Recursent neural network
(2) Mutation testing (3) Code breaking
(4) Filtering Signal processing (5) Learn fuzzy rule base.

* Adv:- (1) Faster (11 processing) (2) Solve problem optim.
* Dis: (1) Fitness func" identification is difficult.
(2) Selection of genetic operator is difficult.

* Procedure:- (1) Generate set of ~~indid~~ individuals.
(2) Genetic operator used, ex: selection or crossover.
(3) Apply mutation or digital reverse if necessary.
(4) Evaluate fitness func" of new population.
(5) Replace predefined pop" with new best individuals.
(6) Repeat 2-5 or terminate if threshold is met.

* Algo with steps:-

Fitness: fitness func", fitness_threshold: termination criterion.
p: No. of hypotheses in population.
r: fraction to be replaced by crossover.
m = mutation rate.
GA(fitness, f" threshold, p, r, m).
(1) Init population: P ← Gen" p hypothesis at random.
(2) Evaluate: for each h in P, comput fitness(h).
(3) While [max_h fitness(h)] < Fitness_threshold, Do:
(i) Select: Probabilitically select (1-r). p members to P to add to P_s.
(ii) Cross over: Prob" select r. p/2 pairs of hypo" from P. For each pair $\langle h_1, h_2 \rangle$ produce 2 child & add to P_s.
(iii) Mutate: Choose m% of members of P_s with uniform prob". For each, invert one randomly selected bit.
(iv) Update: P ← P_s.
(v) Evaluate: Each h ∈ P, compute fitness(h).

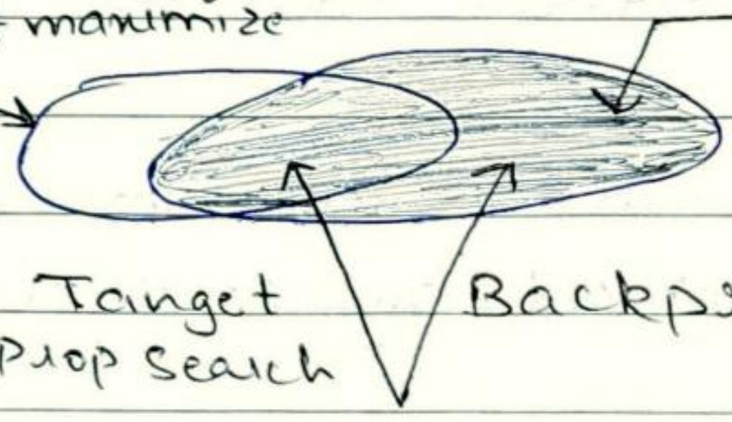
(4) Return hypo" from P that has highest fitness.

* Phases. (6)

(1) Initial population (2) FA fitness func^{n}.
(3) Selection. (4) Crossover.
(5) Mutation. (6) Termination.

→ Hypothesis space search.

* ~~Ste~~ set of possible decision trees.
* Find locally-optimal sol^{n} through hill climbing search.
* ID3 hypo^{n} characterized as searching a space of hypo^{n} for training example fits.



* Maintain single current hypo^{n} i.e candidate elimination.
* No backtracking for converging locally optimal sol^{n}.

→ Genetic Programming: (GP)

* Type of Evolutionary algo (EA), subset of ML.
* Solves problem that humans can't.
* No human interference, sol^{n} better than human.
* Use random mutation, crossover, fitness func^{n}, etc.
* Discover func^{n} relationship b/w features of data.
* Application in SE through code synthesis, genetic improvement, auto bug fixing & etc.

* Adv:- (1) Max length as per hardware limits.
(2) Not necessary that an individual have max knowledge.

* Dis :- (1) Huge No. of possible programs to sol".
(2) very fast but HLL used generate errors & can be slow.
(3) Small variation led to affect solution.

* Types:- (1) Tree based programming.
(2) Stack - based (3) Linear Genetic (LGP)
(4) Grammatical Evolution. (5) Cartesian GP.
(6) Genetic Improvement prog" (GIP).

→ Q. 5.13 for learn-one rule algo.

→ 5.14 for Sequential covering algo.

→ Heuristic Search:- Class of method which is used in order to search a sol" space for an optimal sol" for a problem.

UNIQUE APPSITES

Properties:- (1) Admissibility Condition:- Return opti" sol".
(2) Completeness Condition:- terminate with a sol".
(3) Dominance properties:- Dominate over one member.
(4) Optimality property:- Dominate all member of class.

→ Hill climbing algo:- * local search algo.
* Optimize mathematical problems.
* Also called greedy local search.
* Two component state & value.
* Used when good heuristic is available.
* keeps only single current state.

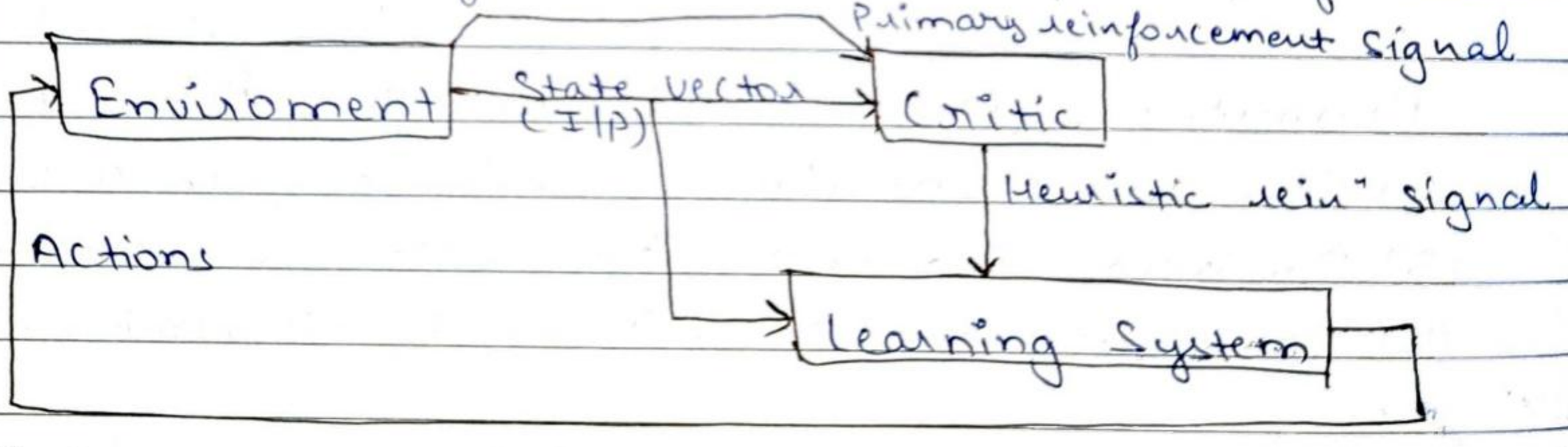
→ Reinforcement learning:-

* How animals & AI can learn to optimize their behaviour in face of rewards & punishments.

* Relates to method of dynamic prog" which is a general approach to optimal control.

* Observed, psychologically, neurobiologically and etc.
* Use optimal policy to max expected total reward.
* Agents learns, what happens when wins (good) and lossed (bad).
* Valuable in robotics. UNIQUE APPSITES
* Robots task of finding out & trial errors.
* learn how to be like humans.
* If successfull rewarded by forward progress & if not penalized by painfull falls.

* +ve & -ve reinforcement are important factors.



Q.5.17 :- Numerical.
Q.5.22 for Diff b/w supervised & Reinforcement.

* Types:- (1) +ve RL (2) -ve RL
* Elements :- (1) Policy (π) (2) Reward func" (r) (3) value func" (V) (4) Transition model (M).

Q3.17,

UAS
Unique AppSites

LIKE
& SUBSCRIBE

DR. A.P.J. ABDUL KALAM TECHNICAL UNIVERSITY, UTTAR PRADESH
LUCKNOW

MACHINE LEARNING
(KOE-073/RCS-080/ROE-083)
AKTU / UPTU
UNIT 5
(GENETIC ALGORITHM)
FULL EXPLANATION

FULL EXPLANATION Unit 5

CLICK HERE

BY ANUBHAV GOEL