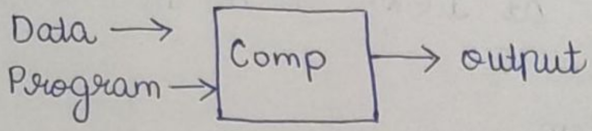
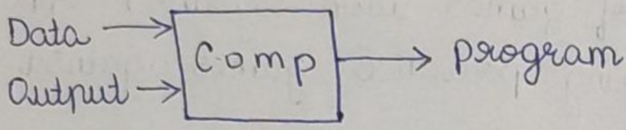
6th September, 2021 Machine Learning L-2

- Traditional learning v/s Machine learning

→ Traditional Programming



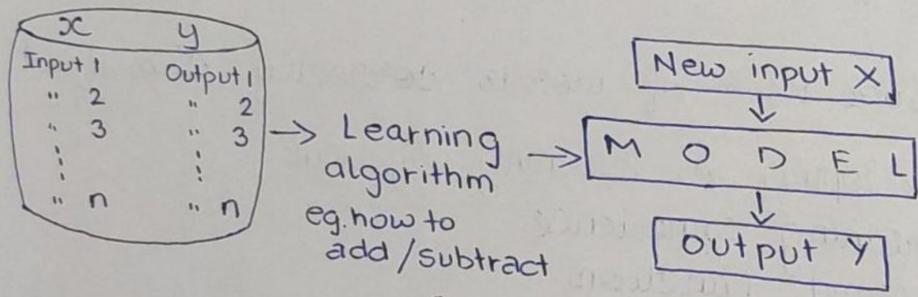
→ Machine Learning



- There are basically 3 types of learning:-

1. SUPERVISED 2. UNSUPERVISED 3. REINFORCEMENT

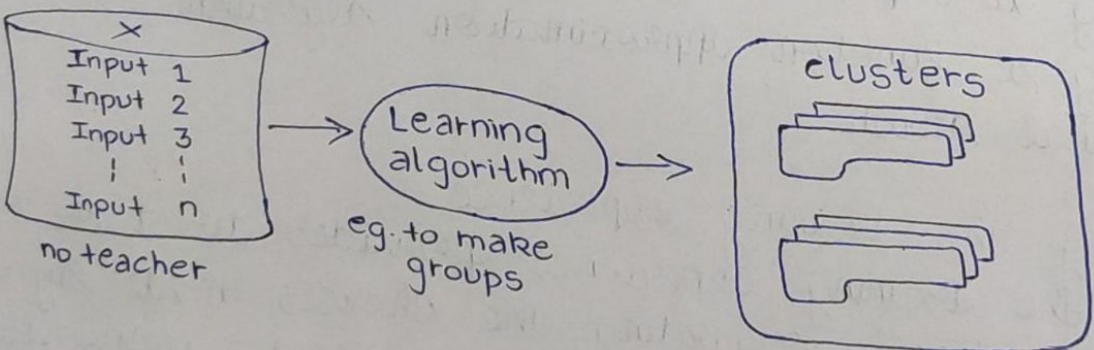
1.



output 1 is like teacher

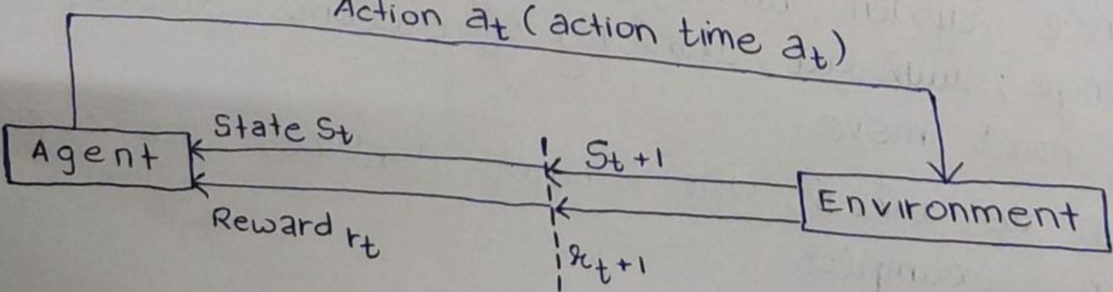
model will be a calculator

2.



a cluster will have same / identical properties

3.



this is a continuous process

□ Well-defined learning problem / Well-posed learning problems :- 1.2

A computer program is said to learn from experience E with respect to some class of tasks T & performance measure P, if its performance at tasks T, as measured by P, improves with experience E.

eg. • A checkers learning problem
- Task T: playing checkers
- Performance measure P: % of games won against oppone.
- Training experience E: playing practice games against itself

◦ A handwriting recognition learning problem
- Task T: recognizing & classifying handwritten words, within images
- Performance measure P: % of words correctly classified

□ Designing & developing a learning system
↳ Choosing the training experience
↳ Choosing the target function
↳ Choosing a representor for target func^n
↳ Choosing a function approximation algorithm
↳ The final design

↳ Choosing the training experience
◦ Whether the training experience provides direct or indirect feedback regarding the choices made by the performance system, with thinking indirect: jeetne tak ka sochenge; without thinking of outcome: direct, only think of next move

The degree to which the learner controls the sequence of training examples.

↳ Choosing the target function
- To determine what tyne of knowledge will be learned, choose move $B \to M$; B (all possible moves) ; M (only valid possibilities)

$V: B \to R$; V is learning function ; B is legal set of moves & R is real value i.e the move that I/ player makes

↳ Choosing a representation for the target function

$V'(b) = W_0 + W_1 x_1 + W_2 x_2 + W_3 x_3 + W_4 x_4 + W_5 x_5 + W_6 x_6$

W_0 = additive constant ; W_1 = numerical coefficient

8.9.21 L-3

↳ Choosing a function approximation algorithm

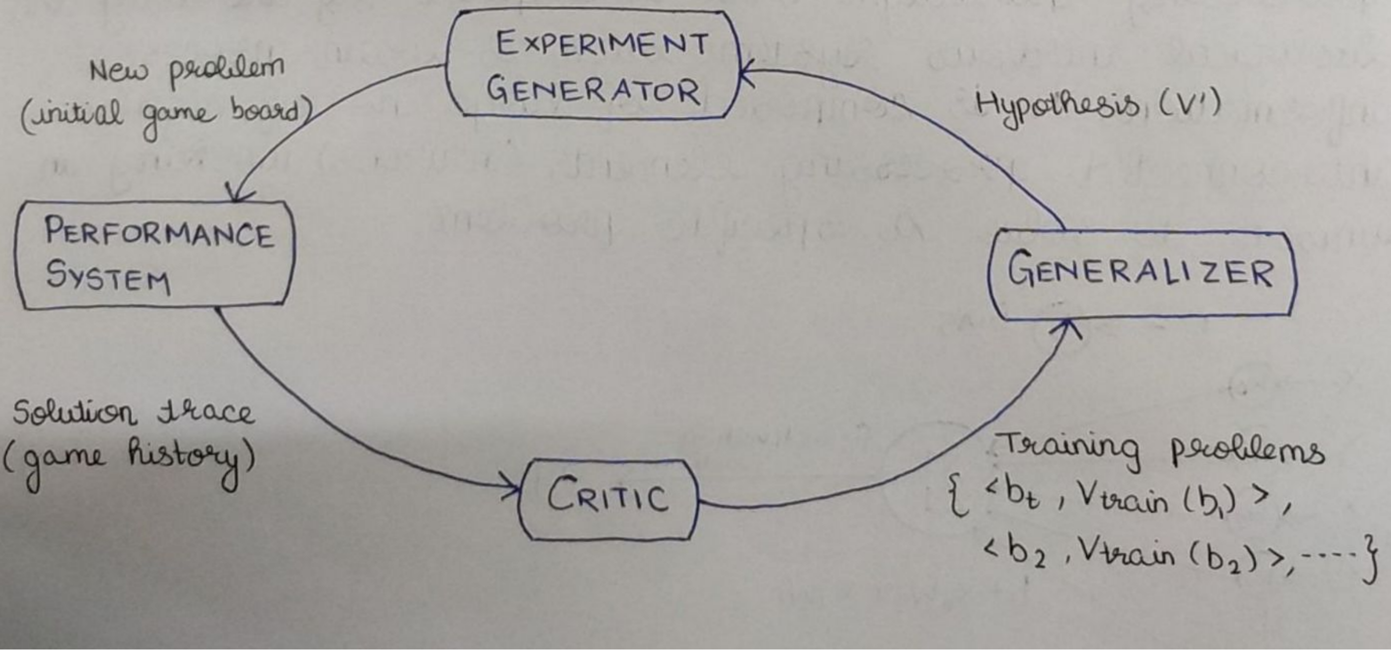
$\underbrace{V_{train}(b)}_{\text{training value}} \longleftarrow V'\text{success}(b)$

$E = \sum (V_{train}(b) - V'(b))^2$; E is squared error

Least Mean Square $W_i = W_i + \eta (V_{train}(b) - V'(b)) x_i$

↑ learning constt. eta

↳ The final design



2.2

i) PERFORMANCE SYSTEM: To solve the given performance task by using the learned target function(s). It takes an instance of a new problem (new game) as input & a trace of its soln (game history) as output.

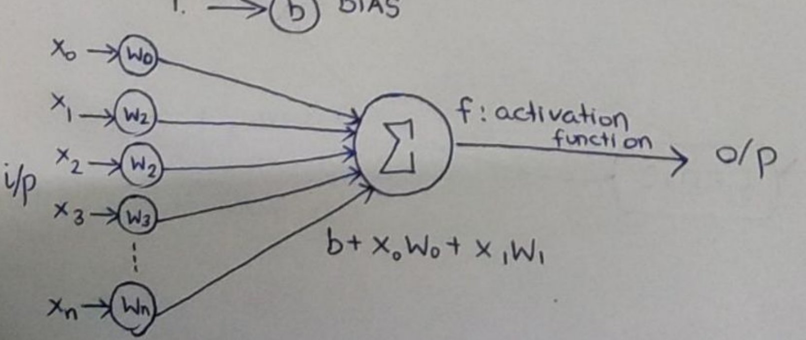
ii) CRITIC: To take as input the history or trace of the game & produce as output a set of training examples of the target func^n.

iii) GENERALIZER: To take as input the training examples & produce as output, hypothesis that is its estimate of the target func^n. It generalizes from the specific training examples, hypothesizing a general func^n that covers those examples & other cases beyond the training examples.

iv) EXPERIMENT GENERATOR: To take as input the current hypothesis (currently learned func^n) & output as new problem (i.e initial board state) for performance system to explore. Its role is to pick new practice problems that will maximize the learning rate of the overall system.

13.9.21 L-4

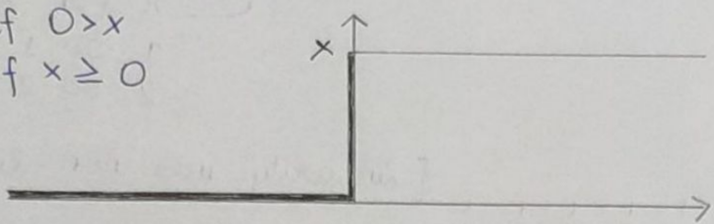
ARTIFICIAL NEURAL NETWORK: ANN is an information processing paradigm that is inspired by the way the biological nervous system such as brain processes information. It is composed of large no. of highly interconnected processing elements (neurons) working in unison to solve a specific problem



□ TYPES OF ACTIVATION FUNCTIONS

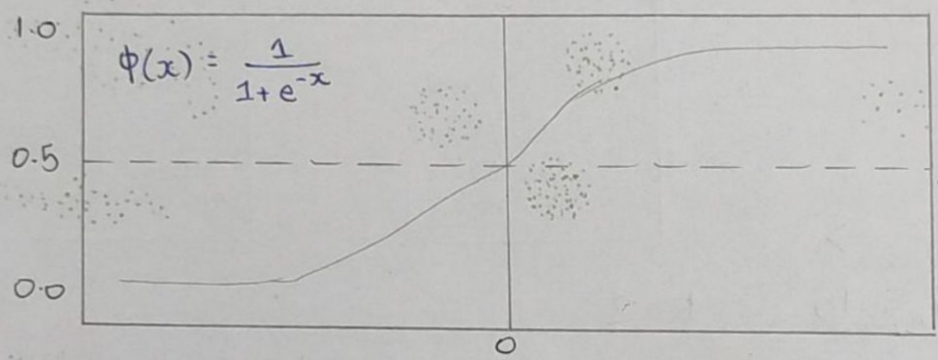
1. Threshold Activation Function

$$f(x): \begin{cases} 0 \text{ if } 0 > x \\ 1 \text{ if } x \geq 0 \end{cases}$$



a binary step func^n

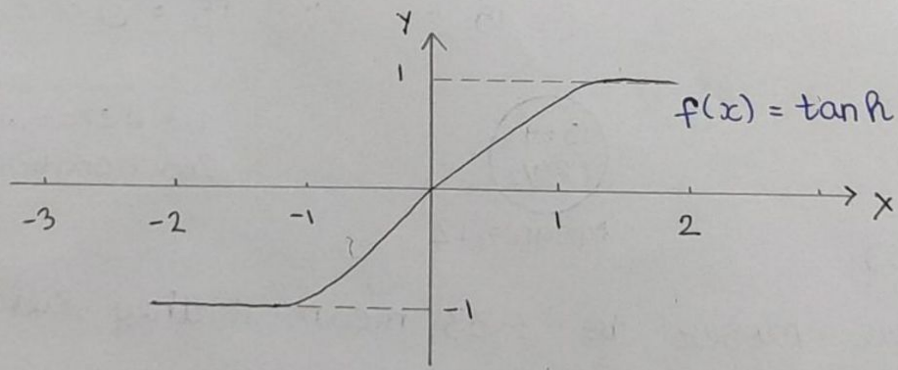
2. Sigmoid Activation Function



$$\phi(x) = \frac{1}{1+e^{-x}}$$

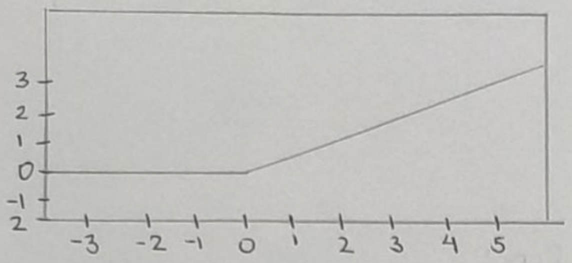
Sigmoid curve logistic function

3. Hyperbolic Tangent Function



$$f(x) = \tanh x$$

4. Rectified Linear Units (ReLu)

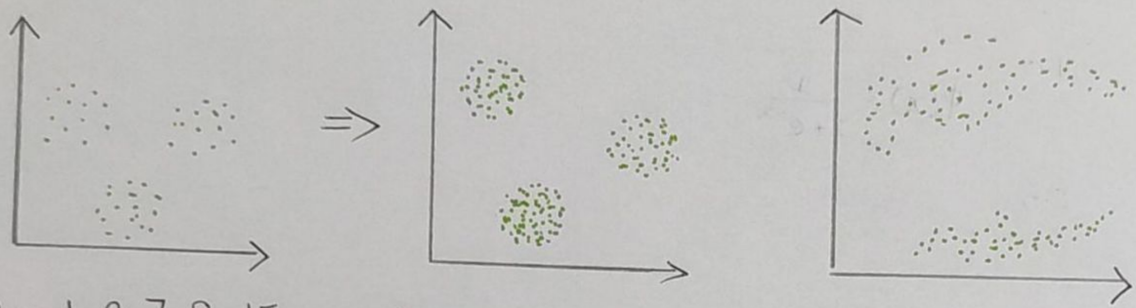


$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x => 0 \end{cases}$$

1 so 1
2 so 2

{ linearity into non-linearity }

CLUSTERING



eg. 1, 2, 7, 9, 15 $\quad$ k = 2

1, 2, 7 $\quad$ 9, 15

1 dist. from 1 = 0
2 dist. from 1 = 1
7 " " 1 = 6
9 " " 15 = 6
15 " " 15 = 0

k is the no. of clusters

7+2+1 = 10/3 $\quad$ 15+9 = 24/2

Mean = 3·33 $\quad$ Mean = 12

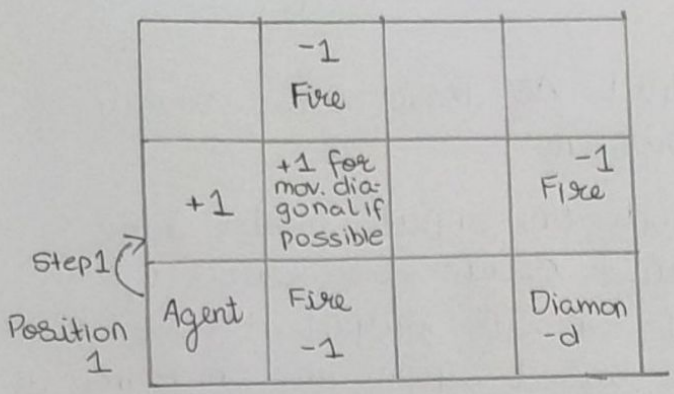
1st iteration, distance
2nd iteration, mean

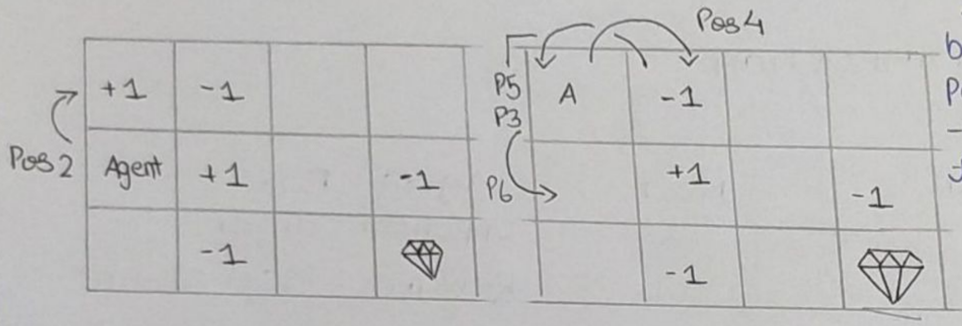
∵ 1,2,7 are closer to 3·33 mean ∴ they belong to that cluster, clustering done

TOPICS FOR NEXT LECTURE:-

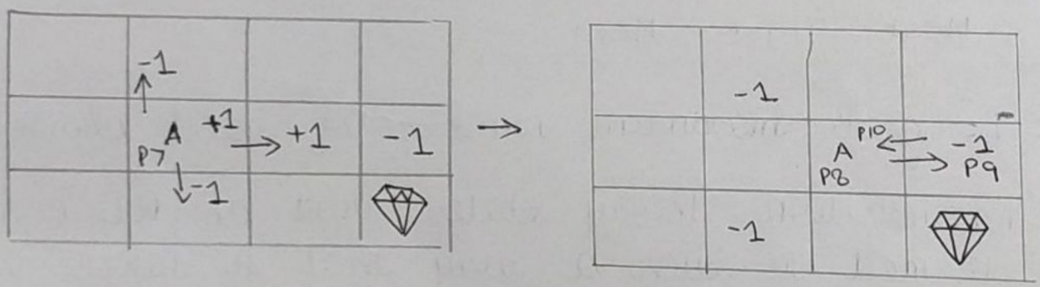
- Reinforcement Learning
- Discounting

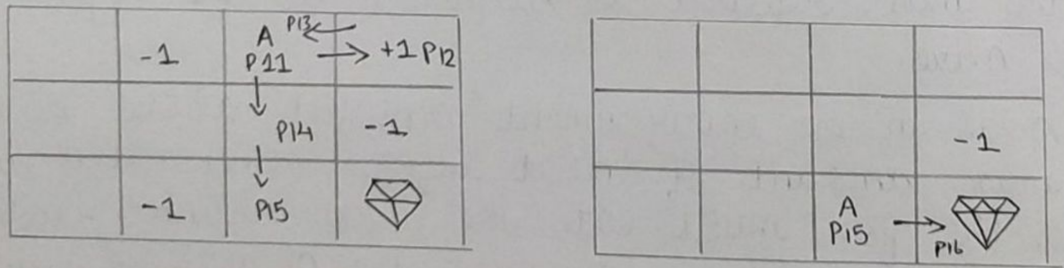
15/9/21 L-5 REINFORCEMENT LEARNING





At Position 3, can't backtrack to initial position so gets -ve reward & then backtracks





AWARD GRANTED

ARTIFICIAL INTELLIGENCE :-

- think logically - think like human
- act logically - act like human

▫ Main points in Reinforcement Learning 4.2

∘ Input : input should be an initial state from which the model will start

∘ Output : many possible output as there are variety of soln to a particular problem

∘ Training : training is based on input model will return a state & user will decide to reward / punish the model based on its output. Model keeps on learning. Best soln is based upon the maximization reward.

▫ REWARD MAXIMIZATION

Agent	Meat		Meat
	Meat	Meat	Meat
	Meat	Tiger	Meat

Agent : Fox
Opponent : Tiger
Reward : Meat chunks

Agent / Machine / Learner

Objective :- To grab maximum number of meat chunks

Reward maximization theory states that an RL agent must be trained in such a way that it takes the best action so that the reward is maximized.

Above, the max. reward is acquired as the agent is still alive.

The end goal in the environment depicted above is to eat the max. amount of meat before being eaten by the tiger. The fox must eat the meat chunks furthest away from the tiger to minimize the chances of being eaten itself. The agent would not eat the chunks closer to the tiger even though they are larger. This is called discounting.

□ Exploitation: involves using the already known information to heighten the rewards.

□ Exploration: means exploring & capturing information about the environment.

So, if fox eats only closer meat chunks to itself, its exploitation; if it goes near tiger for maximum reward its exploration.

□ Introduction to m/c learning approaches

-> Entropy :- degree of randomness

$-p \log_2 p - q \log_2 q$; if $p = 0$, entropy $= 0$

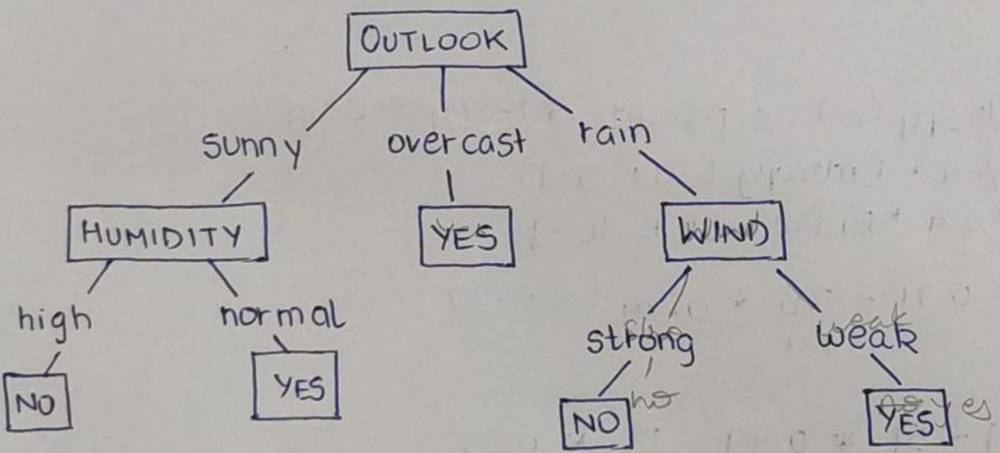
Remember - log base 2 hee use karna!

Decision Tree will come in UNIT - III tennis plays or not

20/9/21 L-6

□ DECISION TREE LEARNING -

max info gain
max entropy

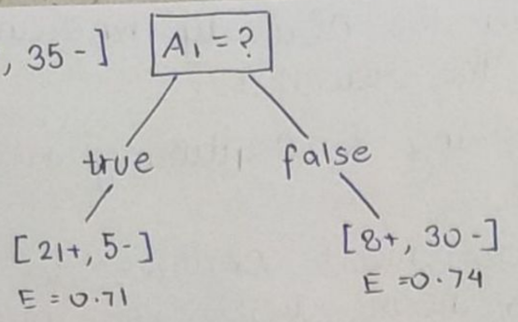


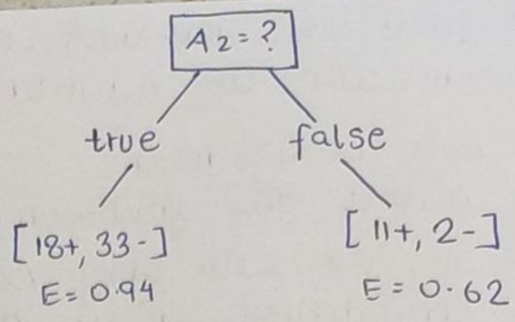
outlook = sunny ∧ humidity = normal

∨ (outlook = overcast)

∨ (outlook = rain ∨ wind = weak)

eg [29+, 35-]





$$-P_+ \log_2 P_+ - P_- \log_2 P_-$$

$$\therefore -29 \log_2 \frac{29}{64} - 35 \log_2 \frac{35}{64} \Rightarrow 0.99 = \text{Entropy}$$

Do your calc.

$$-21 \log_2 \frac{21}{26} - 5 \log_2 \frac{5}{26} =$$

$$-8 \log_2 \frac{8}{38} - 30 \log_2 \frac{30}{38}$$

$$-18 \log_2 \frac{18}{51} - 33 \log_2 \frac{33}{51}$$

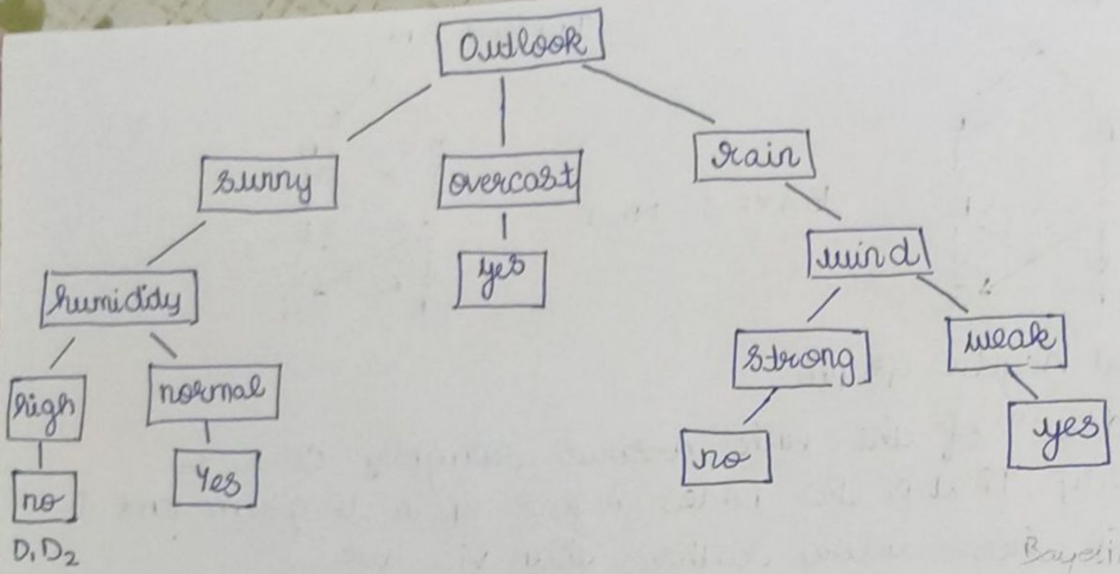
- Gain (S, A) = Entropy (S) i.e parent entropy = 0.99
  - 26/64 * Entropy [21+, 5-])
  - 38/64 * Entropy [8+, 30-]

$$0.99 - \frac{26}{64} * 0.71 - \frac{38}{64} * 0.74 = 0.27$$

- Gain (S, B) = $0.99 - \frac{51}{64} * 0.94 - \frac{13}{64} * 0.62$

  $= 0.12$

→ A provides greater information gain than B.



Bayesian network

22/9/21 L-7

▫ BAYESIAN NETWORKS

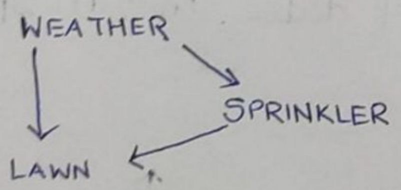
∘ BAYES' RULE :- $P(h|D) = \frac{\overbrace{P(D|h)}^{\text{liklihood}} P(h)}{P(D)}$

understanding Bayes' rule

d = data ; h = hypothesis ; $P(h|D)\,P(d) = P(d|h)\,P(h)$ = joint probability on both sides ; $P(h)$ & $P(D)$ = prior probability

Posterior probability of $h|D$ (↳ dependency) ; h = causal variable ; D = observable variable

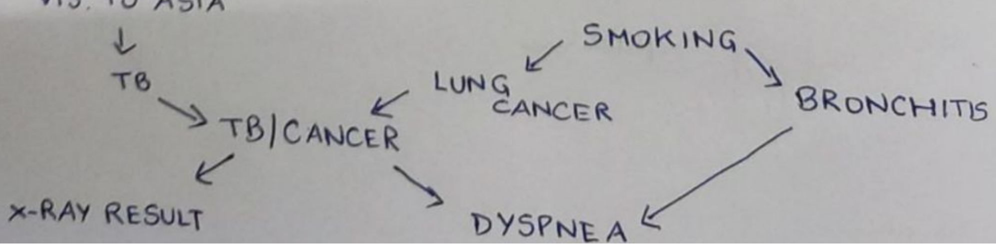
▫ BAYESIAN NETWORK



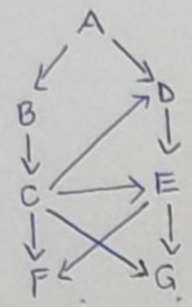
→ loop is allowed
→ cycle isn't "
→ DAG
→ shouldn't backtrack

→ If the lawn is wet, what are the chances it was caused by rain or sprinkler & if the chance of rain increases, how does that affect my having to budget time for watering

Eg of BAY. NET

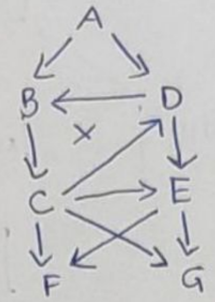


Valid BN



BAYES NET

Invalid BN



Directed Acyclic Graph

→ The direcⁿ of the link arrows roughly corr. to "causality". That is the nodes higher up in diagram tend to influence those below rather than vice versa

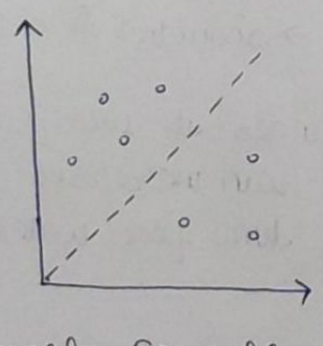
→ In a bayes net, loops may form, cycles shouldn't. This is not an expressive limitation, it does not decrease the power of these nets. (modeling power is not limited)

27/9/21 L-8

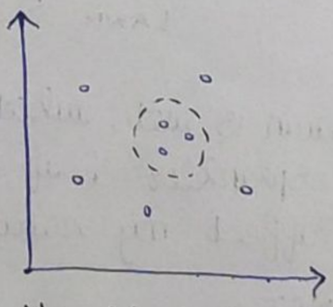
Classification deals with discrete data; regression with continuous

▫ SUPPORT VECTOR MACHINE

→ SVM is a relatively simple Supervised M/C Learning Algo used for classification &/or regression. It's more preferred for Classification but is sometimes very useful for regression too. SVM finds a hyper-plane that crea a boundary b/w types of data. In 2-D space this hyper plane is a line.



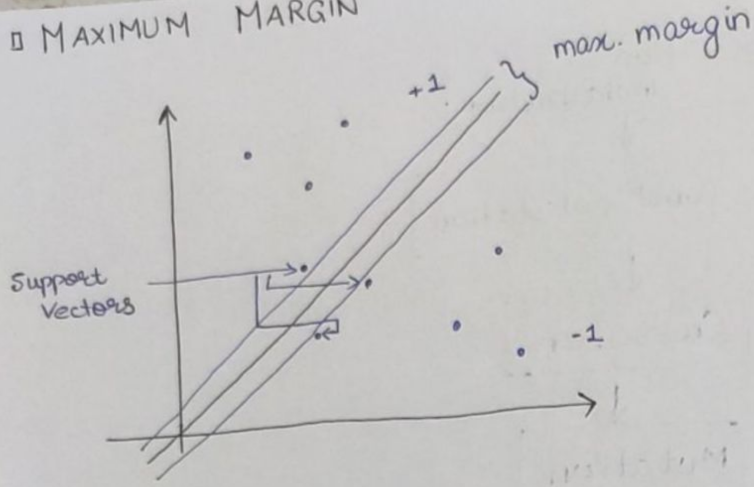
Linearly Separable Data



Non-Linearly Separable Data

→ Kernel

□ MAXIMUM MARGIN



□ GENETIC ALGORITHM

→ chromosome
0101010111011110101100110

↑
gene

Fitness score

A B
0101 1100

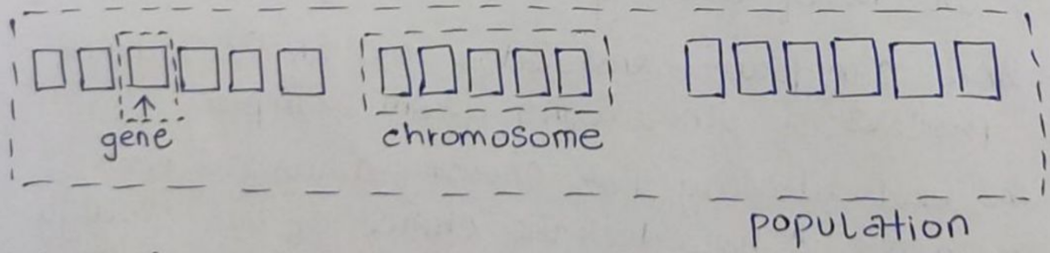
{00 100} child

$(0010)_2 = (2)_{10}$

this is cross-over

Now 1 change to 0 in 0010

∴ $(0000)_2$, if a small change, shows a drastic effect & change, it's called mutation.



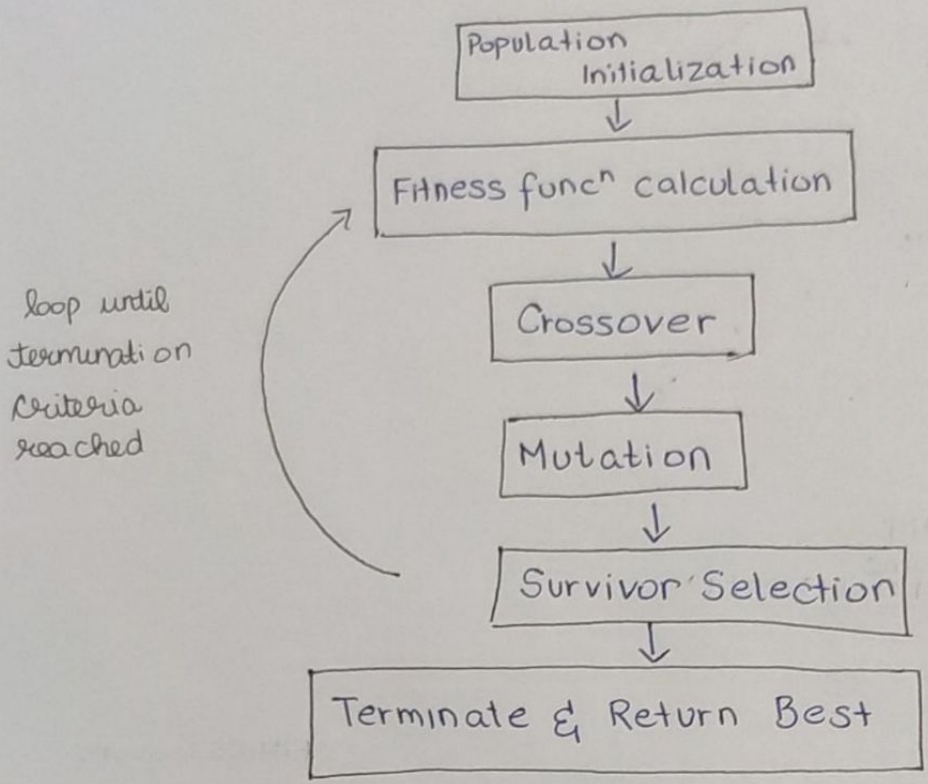
3 types of operation in genetic

→ Selection → Cross over → Mutation

□ Fitness Score

→ F.S is given to each individual which shows the ability of an individual to compete. Individual with optimal or near optimal F.S are sought.

▫ GENETIC ALGO CYCLE



29/9/21 L-8

▫ ISSUES IN MACHINE LEARNING

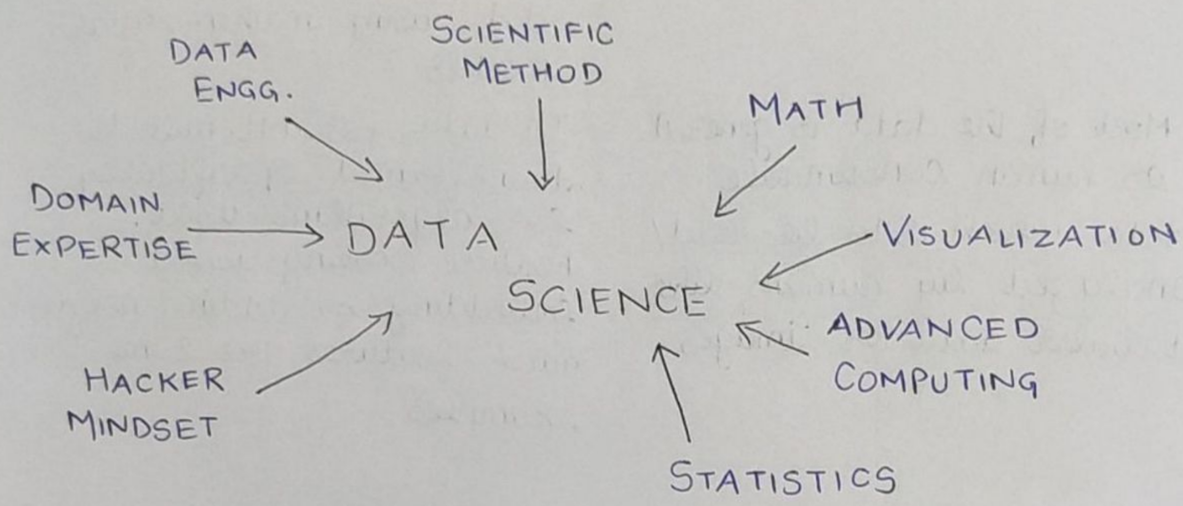
→ Occam' razor : accept the smallest hypothesis & reject the rest

- What algos exist for learning general target func^n from specific training examples
- How does the no. of training examples influence accuracy
- When & how can prior knowledge held by the learner guide the process of generalizing from examples
- What is the best strategy for choosing a useful next training experience, & how does the choice of this strategy alter the complexity of learning problem
- What is the best way to reduce the learning task to one or more func^n approx. problems
- How can the learner automatically alter its repres to improve its ability to represent & learn the target func^n

▫ DATA SCIENCE

-> is an inter-disciplinary field that uses scientific methods, processes, algos & systems to extract knowledge & insights from many structural & unstructured data.

-> DS is related to data mining, machine learning & big data



COMMON DISCIPLINES OF A DATA SCIENTIST

CAPTURE
- data acquisition
- data entry
- signal reception
- data extraction

MAINTAIN
- data warehousing
- data cleansing

PROCESS
- data mining
- clustering / classification
- data modeling
- data summarization

ANALYZE
- exploratory / confirmatory
- predictive analysis
- regression
- text mining
- qualitative analysis

COMMUNICATE
- data reporting
- data visualization
- business intelligence
- decision making

8.2

Data Science VS Machine Learning

SUBJECT	DS	ML
SCOPE	Create insights from data, dealing with all real world complexities	Accurately classify or predict outcomes for new data points by learning patterns from historical data, using mathematical models
INPUT DATA	Most of the data is general. as human consumable data which is to be read/analyzed by humans like tabular data or images	In data for ML will be transformed specifically for algorithms used. Feature scaling, word embedding or adding polynomial features are some examples
SYST. COMPL.	Components for handling unstructured raw data coming.	Major complexity is with
PREF. SKILL SET	Domain expertise, ETL & data profiling, strong SQL, visualization	

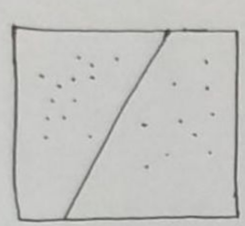
6/10/21 L- LINEAR REGRESSION UNIT-2

□ Classification vs Regression

-> In classification, we use ML algorithms to predict discrete valued outputs

-> In regression we use ML algorithms to predict real valued outputs

discrete value is classification. regression is continuous data range



classification
uses KNN, decision trees & perceptron as ML algorithms

regression
uses linear regression as ML algorithms

□ Regression

-> is a statistical method, that attempts to determine the strength & character of a relationship b/w one dependent variable & a series of other variables (independent variables).

□ Equations/ General Form of Regression

-> Simple linear regression: $Y = a + bX + u$

-> Multiple linear regression: $Y = a + b_1X_1 + b_2X_2 + \cdots b_tX_t + u$

where Y = dependent variable (trying to predict)

X = independent variable (one we are using to predict Y)

a = intercept ; b = slope ; u = regression residual (error)

□ Preliminaries

-> Absolute error : $\Sigma_i [y_i - f(x_i)]$

-> Squared error : $\Sigma_i (y_i - f(x_i))^2$

□ Representation