ML Assignment - II

K·HARIKA ①
ECM-A
21311A1943

Q) Analyze Genetic operators with Examples?

A) * Crossover - It is the operator that performs global exploration ∵ the strings that produced are radically different to both parents in at least some places.

① single point crossover: A position in string is chosen at random & the offspring is made up of first part of parent 1 & second part of parent 2.

② Multi-point: They are chosen, with the offspring being made in the same way.

③ Uniform crossover: Random numbers are used to select which parent to take each element from.

```
1 0 0 1 1 0 0 0 1 0 1
0 0 1 1 1 0 1 0 1 1 0
---------------------
1 0 0 1 1 0 1 0 1 1 0
```

```
Random Samples  0 0 1 1 0 1 1 0 1 1 0
-------------------------------------
string 0        1 0 0 1 1 0 0 0 1 0 1
string 1        0 1 1 1 1 0 1 0 1 1 0
-------------------------------------
                1 0 1 1 1 0 1 0 1 1 1
```

* Mutation:- Genetic operator is mutation, which effectively performs local random search. The value of any element of string can be changed, governed by some probability p.

$$1\,0\,1\,1\,\boxed{1}\,0\,1\,0\,1\,1\,1$$
$$\downarrow$$
$$1\,0\,1\,1\,0\,0\,1\,0\,1\,1\,1$$

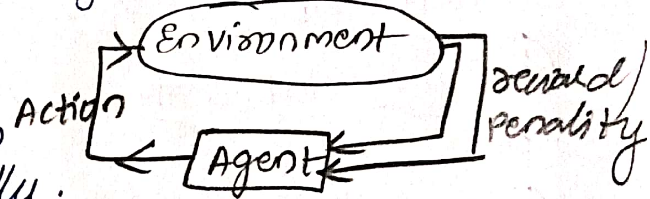
→ Elitism, which takes some number of fittest strings from one generation & puts them directly into next population.

Q) What is Reinforcement learning Explain ?

A) → It is usually described in terms of the interaction between some agent & its Environment.

→ The agent is the thing that is learning, and the Environment is where it is learning, & what it is learning about.

→ For psychological learning theory comes from concept of trial & Error learning, which has been around for long time known "Law of Effect"



→ Maps states or situations to actions in order to maximise some reward numerically.

→ The current sensor readings of robot / processed versions of them, could define the state.

→ Most common way to think about this learning is on a "robot".

→ Robot can drive its motors are the actions, which move the robot in Environment & the reward could be how well it does its task without crashing into things.

Q) What is factor Analysis ?

A) It is used to reduce a large no. of variables into fewer numbers of factors.

→ It is a Statistical data reduction & analysis technique that strives to explain correlations among multiple outcomes as the result of one or more.

→ Techniques Involves data reduction, as it attempts to represent a set of variables by a smaller number

→ It Explicitly assumes the existence of latent factors underlying the observed data.

→

Factor analysis

Exploratory FA

→ It serves as a powerful tool to uncover latent (hidden) factors with in a dataset

Confirmatory FA

→ It is a technique used to evalute pre determined hypothesis (assumptions) regarding the relationship b/w variables & factors.

logics of FA:

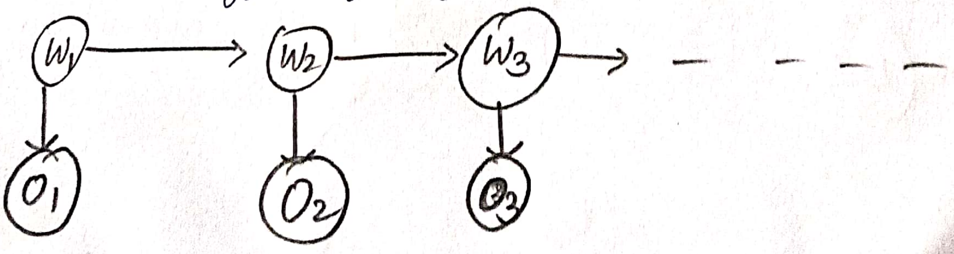
1. Item that you want to reduce (input)
2. Create Mathematical combⁿ to find principle component
3. New combⁿ from residual variables.
4. Select minimal no.of factors
5. Interpret the factors.

Q) Explain HMM with Examples ?

A) → It is a graphical model, used in speech processing & in a lot of statistical work & works on a set of temporal data.

→ Performing inference on the HMM is not computationally expensive, which is a big improvement over the general Bayesian n/w.

→ It is simplest Dynamic Bayesian network, a Bayesian Network deals with sequential data.



- Make an observation of your appearence $O(t)$ & $\omega(t)$ observ... to guess the state.
  - * Probabilities : $P(O_K(t) \mid \omega_j(t))$.
  - * Usually these are labelled as $b_j(O_K)$.
  - * $P(\omega_j(t+1) \mid \omega_i(t))$ and is usually labelled as $a_{i,j}$.

→ 3 things -

① see how well the sequence of observation that I've made match my current HMM

② work out the most probable sequence of states that you've been in based on my observations

③ Given severals sets of observations generate a good HMM for data.

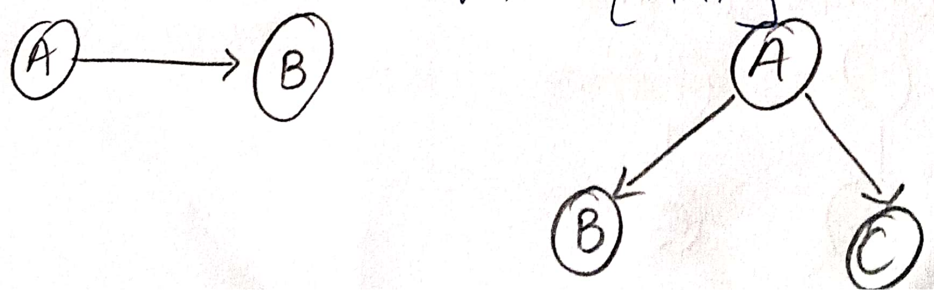
→ HMM Itself is made up of Transistion probabilities $a_{i,j}$ & the Observation probabilities $b_j(O_K)$, & prob of starting in each of state π_I.

Q) Explain Graphical Model?

A) → Uses graph theory with all its underlying computational & mathematical machinery in order to explain probabilistic models.

→ Graphs used in this model are exact ones that are taught in basic algorithms classes: a set of nodes, togethes with links b/w them which can be either directed (or) not.

→ Two types : Directed graphs [Bayesian network]
Undirected graphs [MRF]

→ 

Observ. use directed links bcz these relationships are not symmetrical.
Graph tells us that prob of A & B is same as the prob of A times the prob of B conditioned on A:

$$P(a,b) = P(b|a)\,P(a)$$

→ If there is no directlink b/w 2 nodes then they are conditionally independent of each other.

→ To work out a value for $P(a,b)$, it needs a distribution table for $P(a)$ & one for $P(b|a)$.

→ The nodes are separated into those & their values are directly as observed nodes.

- Bayesian n/w: Directed Acyclic graphs (DAGs), but for graphical models when they are paired with conditional prob tables.
- MRF: It also makes idea of conditional independence that we saw for Bayesian network easier: two nodes in a "Markov Random fields" are conditionally independent of each other.

Q) Write about sampling?

→ Produce samples from prob distributions.
Eg: Initialization of weights.

→ In many cases, the prob distribution used has been the uniform one on [0,1), & done it using Gaussian distributions.

- Random Numbers: Sampling method to generate a random numbers by algorithms that produce pseudo-random numbers is the linear congruential generator.

$$\boxed{x_{n+1} = (a x_n + c) \bmod m}$$

a, c & m are parameters → choose carefully.

→ All the parameters, initial input x_0 (seed) & outputs are integers.

· Gaussian Random Number:

→ Mersenne twister produces uniform random numbers, produce samples from other distributions eg: Gaussian.

→ Usual method is Box-Muller scheme, which uses a pair of uniformly randomly distributed numbers in order to make two independent Gaussian-distributed numbers with zero mean & unit variance.

$$\boxed{f(x,y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \frac{1}{2\pi} e^{-(x^2+y^2)/2}}$$

Q) Differentiate between Analytical vs Inductive learning?

Analytical learning	Inductive learning
Hypothesis fits domain	Hypothesis fits data
Deductive inference	Statistical inference
Learns from scarce data	Requires little prior knowledge
Bias is domain theory	Syntactic inductive bias.
Perfect prior knowledge & scarce data	Plentiful data, No prior knowledge.

Explain

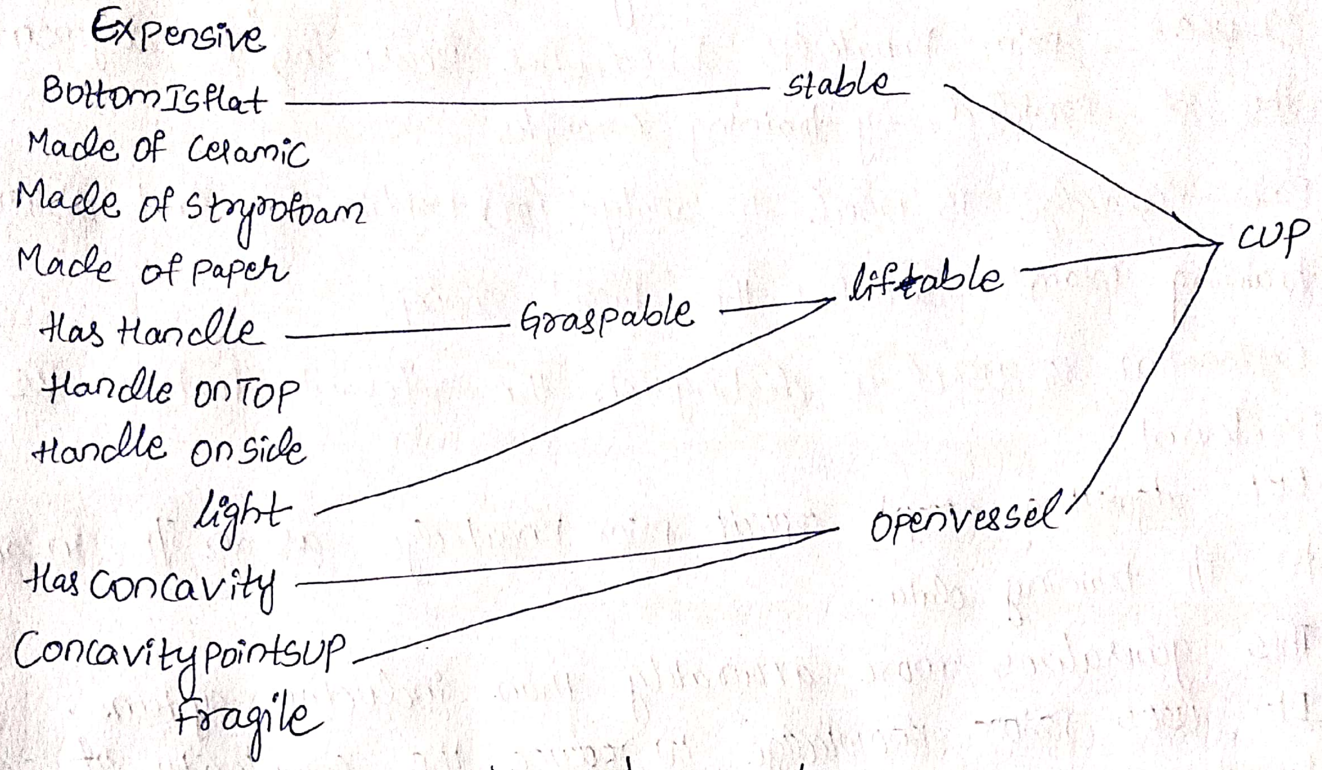
KP

Explain KBANN Algorithm ?

KBANN (data D, domain Theory B)

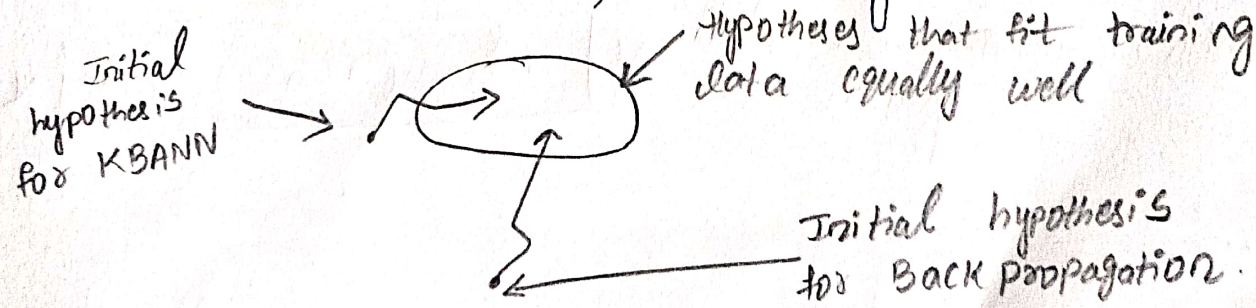
1. Create a feedforward network h equivalent to B

2. Use BACKPROP to tune h to fit D.



- Create one unit per horn clause rule
  - → Connect unit inputs to corresponding clause antecedents
  - → For each non-negated antecedent, corresponding input weight $w \leftarrow W$, W = constant
  - → $w \leftarrow -W$ (negated antecedent, input weight)
  - → Threshold weight $w_0 \leftarrow -(n - .5)W$, where n is no. of non-negated antecedents.

Liftable ← Graspable, ¬Heavy.



Q) Analyze Explanation Based learning (EBL)?

A) → Inductive learning methods such as Neural network & Decision Tree learning, Inductive logic programming & Genetic Algorithms. requries a certain no. of training Examples to achieve a given level of generalization accuracy.

→ It uses 'Prior Knowledge & Deductive Reasoning" to augment the inf provided by training Examples

→ Prior Knowledge is used to analyze (or) Explain, how Each observed training Example satisfies the target concept.

→ Explanation is used to distinguish the relevant features vs irrelevant.

→ EBL algorithm accept Explicit prior knowledge as an i/P, in add^n to i/P training data.

→ This generalizes more accurately than inductive systems.

→ It uses prior knowledge to reduce the complexity of hypothesis space to be searched.

Eg:- task of learning to play chess.