Saturday
13-6-2020

AI Assignment - 4
unit - 4

N.V.S.k.kalyani
174N1A0584
3rd B.tech CSE-C

2 Marks

1A) i, Random variables & Types :-

The Basic element of the language is the random variable, which canbe thought of as referring to a "part" of the world whose "status" is Intially unknown.

Types :- i, Boolean random variables
ii, Discrete random variables
iii, continous random variables

ii, Atomic event :-

The notion of an atomic event is useful in understanding the foundations of probability theory.

An atomic event is a complete specification of the state of the world about which the agent is uncertain.

iii, prior probaility :-

The unconditional or prior probability associated with proposition α is the degree of belief accorded to in the absence of any other information. It is written as p(a).

iv, posterior probability :-

once the agent has obtained some evidence concering the previously unknown random variables making up the domain, prior probabilities are no longer applicable.

v, Join probability :- $p(A \text{ and } B) = p(A \wedge B) = p(A) * p(B)$

2A) **Axioms of Probability:-**

- let S be a sample space, A and B are events

$P(A) > 0$
$P(S) = 1$
$P(A') = 1 - P(A)$
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

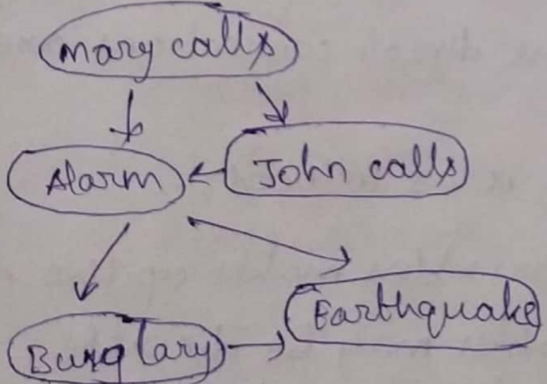
This is called addition Rule of probability

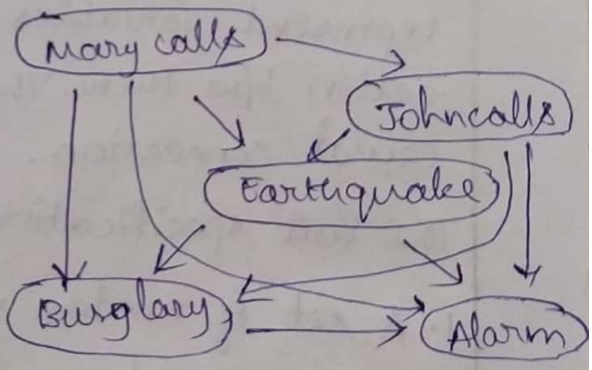
=) If events A and B are mutually exclusive, then $P(A \cup B) = P(A) + P(B)$.

3A) **Bayesian belief network:-**

Bayesian networks are graphical models for reasoning under uncertainly, where the node represents variable and arcs represent direct connections b/w them. These direct conn-ections are often casual connections.

4A)





5A) function FORWARD_BACKWARD(ev, prior) returns a vector of probability distribution

Inputs: ev, a vector of evidence value for steps 1...t prior, the prior distribution on the initial state, $P(x_0)$

local variables: fv, a vector of forward messages for step so,...t, b, a representation of the backward message initially all 1s

sv, a vector of smoothed estimates for steps 1,....t

```
fv[0] ← prior
for i=1 to t do
fv[i] ← forward[fv[i-1], ev[i])
for i=t downto 1 do
sv[i] ← NORMALIZE(fv[i] x b)
b ← BACKWARD (b, ev[i)]
return sv
```

Essays :-

1A) **Definition** :- Bayesion Networks are graphical model for reasoning under uncertainty, where the nodes represent variables and arcs represent direct connection b/w them. These direct connections are often casual connection.

The full specification is as follows:

1. A set of random variables makes up the nodes of the network variables may be discrete.
2. A set of directed links or arrows conceeds pairs of nodes. If there is an arrow from node x to node y, x is said to be a parent of y.
3. Each node x, has a conditional probability distribution p(x) (parents(x)) that qualifies the

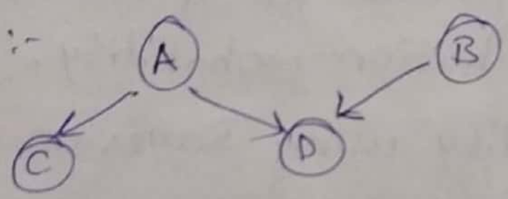
effect of the parents on the node.

4. The graph has no, directed cycle.

* A topology of the network. The set of nodes and links specifies the conditional Independence relationships the hold in the domain, in a way that will be made precise shortly.

* The Intuitive meaning of an arrow in a properly construct network is usually that X has a direct Influence of Y.

Example:-



* To describe above Bayesian Network, we should specify the following probabilities.

$P(A) = 0.3$

$P(B) = 0.6$

$P(C/A) = 0.4$

$P(C/\sim A) = 0.2$

$P(D/A.B) = 0.7$

$P(D/A\sim B) = 0.4$

$P(D/\sim A, B) = 0.2$

$P(D/\sim A, \sim B) = 0.01$

* They can also be expressed as conditional probability tables as follows:

17UN1A0584

conditional probability Tables						
P(A)	P(B)	A	P(C)	A	B	P(D)
0.3	0.6	F	0.2	T	T	0.7
		T	0.4	T	F	0.4
				F	T	0.2
				F	F	0.01

* using Bayesion belief network given, only 8 probability values in contrast to 16 values are required in general for 4 variables {A,B,C,D} in joint distribution probability.
* joint probability using Bayesion Belief network is computed as follows:

$$P(A,B,C,D) = P(D/A,B) * (P(C/A)) * P(B) * P(A)$$

$$= 0.7 * 0.4 * 0.6 * 0.3$$

$$= 0.0504$$

conditional probability Table (CPT) :-

* It contains conditional distribution.
* This form of table can be used for discode val also suitable for continous variables.
* Each row in a CPT contains the conditional probability value for a conditioning case.
* A conditioning case is just a possible combi-nation of values for the parent node a miniat atomic event, if you like.
* Each row must sum to 1, because the entries represent an exhaustic set of cases for the

Semantics of Bayesian Network:-

* There are two ways in which one can under-stand the semantics of Bayesian network.
* The first is to see the network as a represent-ation of the joint probability distribution.
* The second is to view it as to encoding of a collection of conditional independence statement
* The two views are equivalent.

prepresenting full joint distribution:

A generic entry in the joint distribution is the probability of a conjunction of particular assignments to each variable, such as

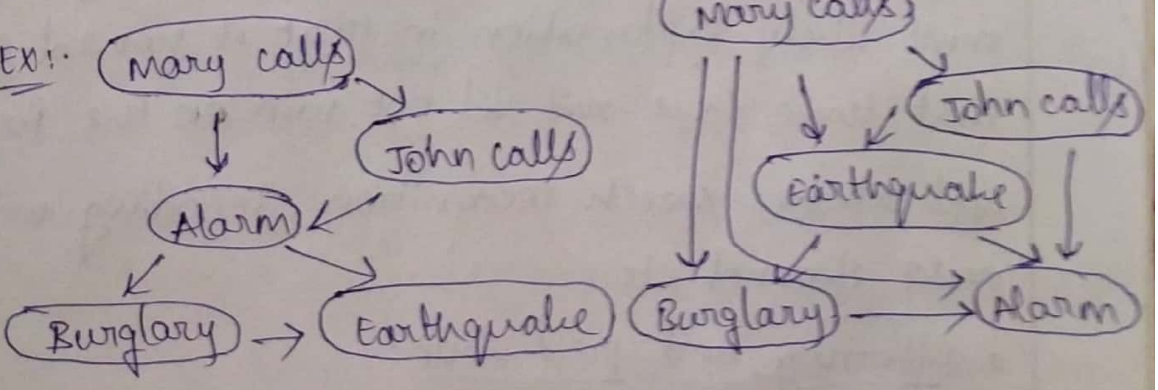
$$P(x_1 = x_1 A, \ldots A x_n = x_n).$$

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{N} p(x_i | parent\_node(x_i)) \rightarrow \text{Equation } 1$$

compactness and node ordering:-

* The compactness of Bayesian network is an Ex of a very general property of locally structured (also called sparse) system.

Ex:-



1T4N1A0584

2A) Temporal models

Having set up the structure of a generic tempo[ral] method. we can formulate the basic inference task that must be solved.

i, Filtering : $p(x_t/e_{1:t})$

To compute the current belief state given all evidence. Better name : state estimation.

ii, prediction : $p(x_{t+k}|e_{1:t})$ for $k>0$

To compute a future belief state, given current evidence.

Ex:- In the umbrella, this might mean computing the probability of rain three days from now.

iii, smoothing : $p(x_k|e_{1:t})$ for $0\leq k<t$. To compute a better estimate of past states.

iv, most likely Explanation : $\arg\max_{x_{1:t}} p(x_{1:t}|e_{1:t})$

To compute the state sequence that is most likely given the evidence.

Ex:- If the umbrella appears on each of the first three days and is atleast on the fourth, then the most likely explanation is that it rained on the first three days and did not rain on the fourth.

Applications:- speech recognition, decoding with a noisy channel etc.,

i, Filtering and prediction:-

* A useful filtering algorithm needs to maintain a current state estimate and update it, rather

than going back over the entire history of percepts of each update.

* In other words, given the results filtering upto time the agent needs to compute the results for the t+1 from the new evidence e_{t+1}.

$$P(x_{t+1} \mid e_{i:t+1}) = f(e_{t+1}, P(x_t \mid e_{1:k}))$$ recursive estimation

* It can be two-parts process emerge quite simply when the formula is rearranged.

$$p(x_{t+1} \mid e_{1:t+1}) = p(x_{t+1} \mid e_{1:t}, e_{t+1})$$ (diving up the evidence)

$$= \alpha p(e_{t+1} \mid x_{t+1}, e_{1:t}) P(x_{t+1} \mid e_{1:t})$$ (using Bayes rule)

$$= \alpha p(e_{t+1} \mid x_{t+1}) P(x_{t+1} \mid e_{1:t})$$ (sensor markov assumption)

* The second term, $p(x_{t+1} \mid e_{1:t})$ represents a one-step productions predictions of the next state, and the first term, updates this will the new evidence. Notify that $p(e_{t+1} \mid x_{t+1})$ is obtainable directly from the sensor model.

* Now we obtain the one step prediction for the next state by conditioning on the current state

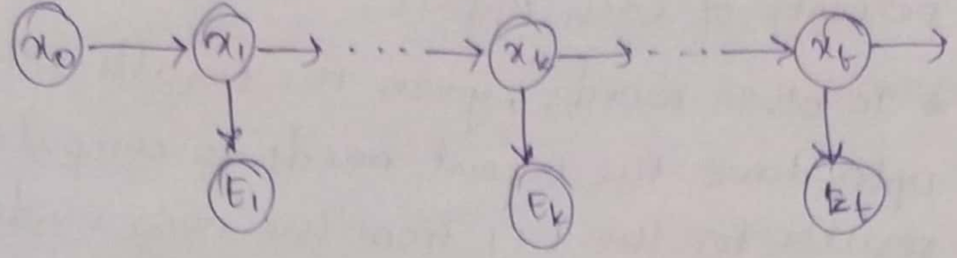
$$x_t = p(x_{t+1} \mid e_{1:t+1}) = \alpha p(e_{t+1} \mid x_{t+1}) \sum_{x_t} P(x_{t+1} \mid x_t, e_{1:t}) P(x_t \mid e_{1:t})$$

$$= \alpha p(e_{t+1} \mid x_{t+1}) \sum p(x_{t+1} \mid x_t) p(x_t \mid e_{1:t})$$

ii) Smoothing:- It is a process of computing msgs passing approach we can split the computation into two parts - the evidence up to 'k' and the

17UN1A0584

evidence from k+1 to t.



proof:- $P(x_k|e_{1:t}) = P(x_k|e_{1:k}, e_{k+1:t})$

$= \alpha P(x_k|e_{1:k}) P(e_{k+1:t}|x_k, e_{1:k})$

$= \alpha P(x_k|e_{1:k}) P(e_{k+1:t}|x_k)$

$= \alpha f_{1:k} \times b_{k+1:t}$

* It terms out that the backward message $b_{k+1:t}$ can be computed by a recursive process. that runs backward from: t

$$P(e_{k+1:t}|x_k) = \sum_{x_{k+1}} P(e_{k+1:t}|x_k, x_{k+1}) P(x_{k+1}|x_k)$$

$$= \sum_{x_{k+1}} P(e_{k+1:t}|x_{k+1}) P(x_{k+1}|x_k)$$

$$= \sum_{x_{k+1}} P(e_{k+1}, e_{k+2:k}|x_{k+1}) P(x_{k+1}|x_k)$$

$$= \sum_{x_{k+1}} P(e_{k+1}|x_{k+1}) P(e_{k+2t}|x_{k+1}) P(x_{k+1}|x_k)$$

Forward and Backward Algorithm:-

function FORWARD-BACKWARD (ev. prior) returns a vector of probability distribution.

inputs: ev. a vector of evidence values for 1,... t

prior, the prior distribution on the initial state $p(x_0)$.

17UN1A0584

local variables : fu, a vector of forward message for steps 0,...t. b, a representation of the backward message, intially all is.
sv, a vector of smoothed estimates for steps 1,...t.

fu[0] ← prior
for i=1 to t do
fu[i] ← FORWARD (fu[i-1], ev[i])
for i=t downto 1 do
su[i] ← NORMALIZE (fu[i] x b)
b ← BACKWARD (b, ev[i])
return sv

iii, Finding the most likely sentence :-

* Suppose that [true, true, false, true, true] is the umbrella sequence for the security cards first five days on the Job.
* what is the weather sequence must likely to explain this?
* Does the absence of the umberella on day 3 mean that it wasn't raining or did the director forget to bring it?
* Thus the algorithm for computing the most likely sequence is similar to filtering it runs - forward along the sequence, compose

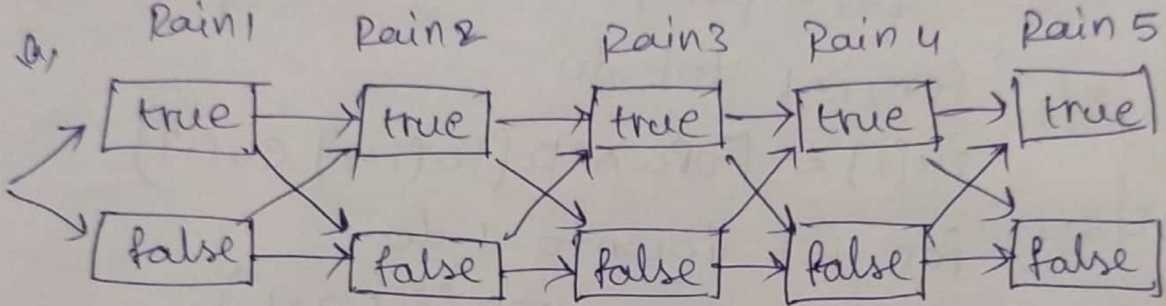
$$\max_{x_1 \ldots x_t} P(x_1, \ldots, x_t, x_{t+1} \mid e_{1:t+1})$$

174N1A0584

$$= \alpha\, p(e_{t+1} | x_{t+1}) \max_{x_t} \left(P(x_{t+1} | x_t) \max_{x_1 \ldots x_{t-1}} P(x_1, \ldots, x_{t-1}, x_t | e_{1:t}) \right)$$

The progress of computation is shown in Fig. At the end, it will have the probability for the most likely sequence reading each of the final state.

a,



b,