DATE
PAGE

Dhananjay Porwal (EN19CS301110)

CS-B

Pattern Recognition

DATE
PAGE

Assignment 1

1 Explain Bayesian Belief Network?

Ans Bayesian belief networks provide an intermediate approach which allows stating conditional independence assumptions that apply to subsets of the variables

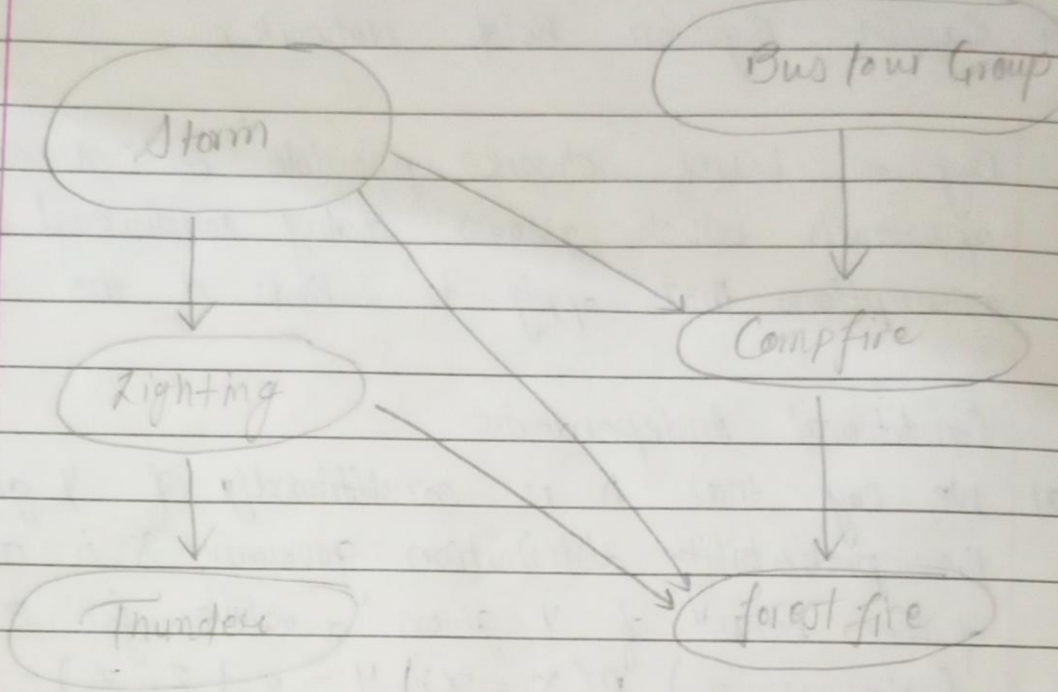
Conditional Independence

a) We say that X is conditionally of Y given Z if the probability distribution governing X is independent of the value of Y given a value for Z. ie
$(\forall x_i, y_j, z_k)\ P(X = x_i | Y = y_j | Z = z_k) = P(X = x_i | Z = z_k)$
or $P(X | Y, Z) = P(X | Z)$

b) This definition can be extended to sets of variables as well as we say that the set of variables $X_1 \ldots X_l$ is conditionally independent of the set of variables $Z_1, Z_2 \ldots Z_n$, if $P(X_1 \ldots X_l | Y_1 \ldots Y_m, Z_1 \ldots Z_n) = P(X_1 \ldots X_l | Z_1 \ldots Z_n)$

Representation in Bayesian Belief Networks
Associated with each node is a conditional probability table, which specifies the conditional distribution for the variables gives its imediate parents in the

graph.



Each node is asserted to be conditionally independent of its non-descendants, given its intermediate parents.

Learning Bayesian Belief Networks -

Case I -

The network structure is given in advance & all the variables are fully observable in the training. Ex - ==> Trivial, just estimate the conditional probabilities.

DATE
PAGE

Case II –

The network structure is given in advance but only some of the variables are observed in the training of data. for ex – ==> Similar to learning the weights for the hidden units of a Neural Net : Gradient the Ascent procedure.

Case III –

The network structure is not known in advance. ==> Use a heuristic search or constraint based technique to search through potential structures.

$$P(a_3, b_1, x_2, c_3, d_2) = P(a_3) P(b_1) P(x_2 / a_3, b_1) P(c_3 / x_2) P(d_2 / x_2)$$

$$= 0.25 \times 0.6 \times 0.4 \times 0.5 \times 0.4$$

$$= 0.012$$

2 Explain maximum likelihood estimation with ex?

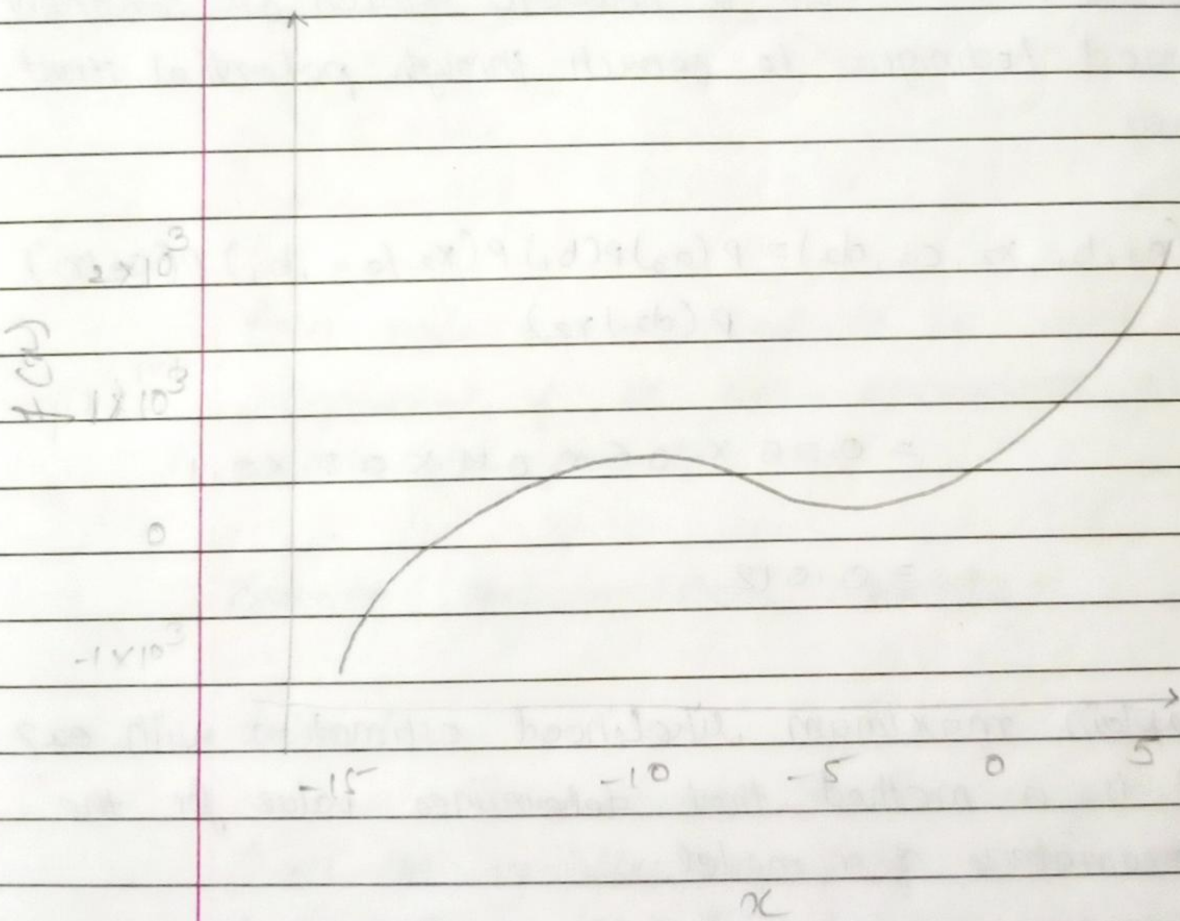
Ans It is a method that determines value for the parameters of a model.

$$P(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

In our example the total (joint) probabilities density of observing the three data points is given by

$$P(9, 9.5, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(9-\mu)^2}{2\sigma^2}\right) \times \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(9.5-\mu)^2}{2\sigma^2}\right)$$

$$\times \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(11-\mu)^2}{2\sigma^2}\right)$$



DATE
PAGE

This is the example of non-monotonic function because as you go from left to right on the graph of $f(x)$ goes up, then goes down & then goes up again. Taking logs of the original expression given as -

$$\ln(p(x,\mu,\sigma)) = \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(9-\mu)^2}{2\sigma^2} + \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(9.5-\mu)^2}{2\sigma^2} +$$

$$\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(11-\mu)^2}{2\sigma^2}$$

Thus,

$$\mu = \frac{9+9.5+11}{3} = 9.833$$

Thus we have our maximum likelihood estimation for μ. We can do the same things with σ too.

Q Explain Hidden Markov Models (HMMs)?

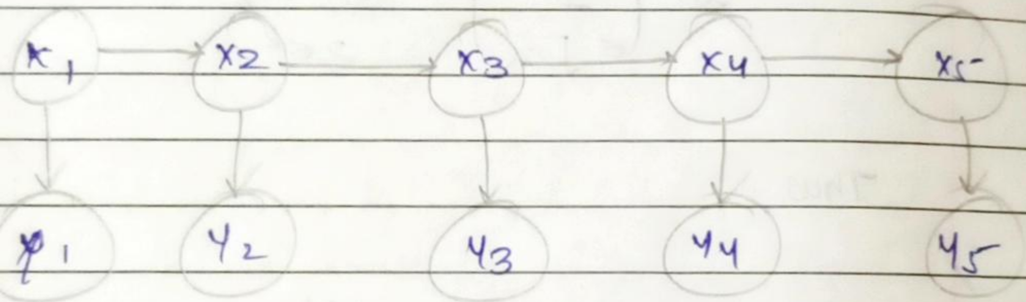
Ans Hidden Markov models (HMMs) are a type of statistical modeling that has been used for several years. They have been applied in different fields such as medicine, computer science, & data science.

Hidden Markov Models

$p(y_t \mid x_t)$ Observation probability SONAR noisiness

$p(x_t \mid x_{t-1})$ Transition probability Submarine locomotion

$$P(x, y) = p(x_1) \prod_{t=1}^{T-1} p(x_{t+1} \mid x_t) \prod_{t=1}^{T} p(y_t \mid x_t)$$



4 Explain Discriminant function?

Ans A function of a set of variables that is evaluated for samples of events or objects & used as an aid in discriminating between or classifying them.

The principle is the same as the two-category cases -

Decision Rule

if $P(w_i|x) = \max\limits_{j=1,2...c} P(w_j|x)$, then $x \in w$

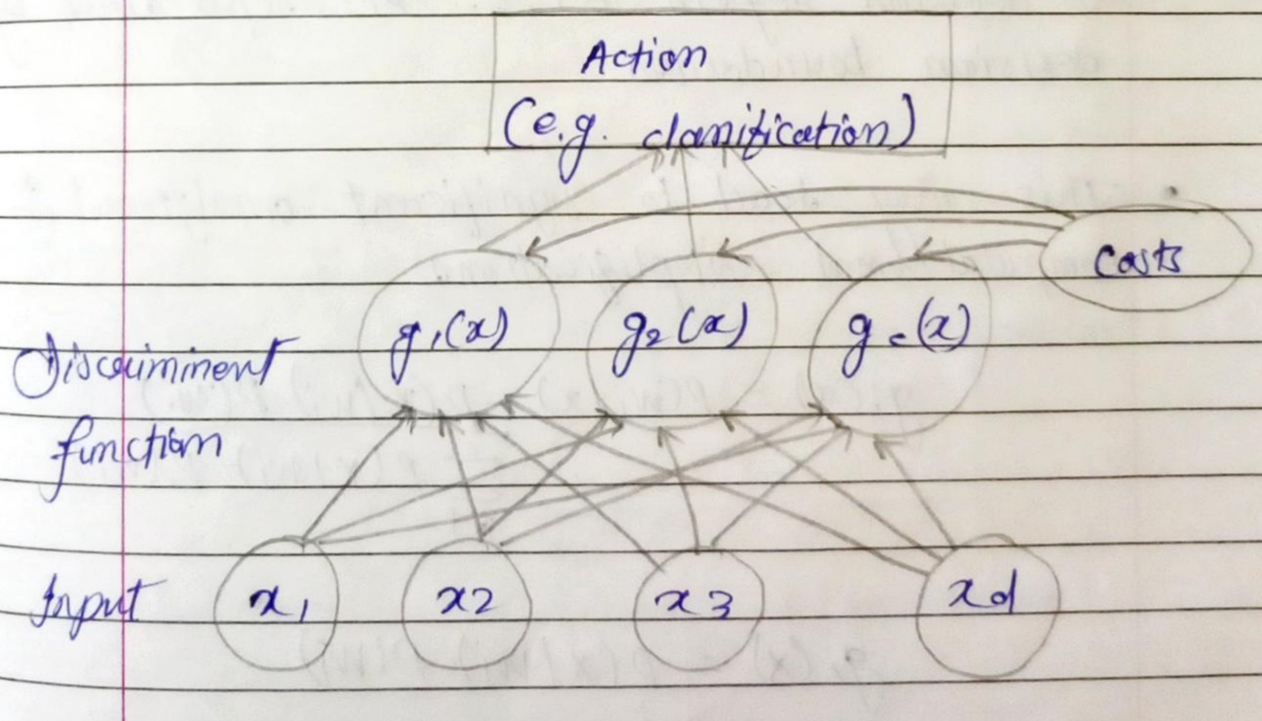
$\uparrow$

Or: Discriminant function $g_i(x)$

$\downarrow$

if $P(x_i|w_i)\,P(w_i) = \max\limits_{j=1,2...c} P(x|w_j)P(w_j)$, then $x \in w$

Decision is made by comparing the discrimination functions of each class.



DATE
PAGE

- A useful way of representing classifiers is through discriminant function $g_i(x)$, $i=1,\dots,c$ where the classifier assigns a feature vector x to class w_i, if

$$g_i(x) > g_j(x) \quad \forall j \neq i$$

$$g_i(x) = -R(\alpha_i | x)$$

$$g_i(x) = P(w_i | x)$$

- This functions divide the features space into c decision region $(R_1, \dots R_c)$, separated by decision boundaries

- This may lead to significant analytical & computational simplifications

$$g_i(x) = P(w_i | x) = \frac{p(x|w_i)\, P(w_i)}{\sum_{j=1}^{c} p(x|w_j)\, P(w_j)}$$

$$g_i(x) = p(x|w_i)\, P(w_i)$$

DATE
PAGE

$$g_i(x) = \ln p(x|w_i) + \ln P(w_i)$$

$$g(x) = g_1(x) - g_2(x)$$

5 What is pattern recognition? Differentiate between supervised & unsupervised learning?

Ans Pattern Recognition assign an unknown pattern to one of several known categories.

The act of taking raw data & taking an action based on the "category" of the pattern. we gain an understanding & appreciation for pattern recognition in the real world - visual scene. noises, etc.

Human senses - sight, Hearing, Taste, Smell, Touch.

A pattern could be an object or event.

There are 3 types of PR -

a) STASTICAL PR (Decision Theorem)

b) SYNTATICAL PR (structure)

c) NEURAL PR

Application of pattern recognization -

a) Image processing.

b) Fingerprint Identification.

c) Charactor Recognition.
etc...

Supervised Learning -
Where you have input variables (x) & output variables (y) & you use an algorithm to learn the mapping function from the input to the output.

$$y = f(x)$$

The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variable (y) for that data it is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervisor to the learning process.

a) Classification (Binary decisions)

b) Regression (Dependency between data)

Unsupervised Learning -

Where you only have input data (x) & no corresponding output values.

The goal for this learning is to model the underlying structure or distribution in the data in order to learn more about data this is called as unsupervised Learning.

a) Clustering.

b) Association.