MONISH KUMAR . E . M

2023115076

IT - 1

ML ASSIGNMENT - 1

Unit-4 Probabilistic graphical models.

1) Naive Bayes Algorithm:-

The Naive Bayes algorithm is a classification technique based on bayes theorem. It assumes that all features are conditionally independent given class label

$$P(C|X) = \frac{P(X|C)\ P(C)}{P(X)}$$

where,

C → class

X → feature vector

Since P(X) is constant the decision is based on maximizing P(X|C) P(C)

Applications:-

* Spam detection
* Medical diagnosis
* Sentiment analysis.

2) Bayesian belief network (BBN):-

A bayesian belief network (BBN) also known as a bayesian network, is a graphical model that represents probablistic relationships among a set of random variables.

It is represented as a directed acyclic graph (DAG).

=) Nodes represent random variables.

-) Edges represents conditional dependicies.

-) Each node has conditional probability table (CPT) that quantifies effect of parent nodes.

example:-

consider medical diagnosis network:

Node1 : flue

Node2: fever

Node3: Cough

Here fever and cough depends on flue.

The network completely encodes there dependencies.

Advantages:

=) Captures causal relationship.

=) Handles incomplete data

=) Support both inference (predict unknown) and learning (update probabilities)

Application:-

=) Medical diagnosis (disease & symptoms)

=) Fault detection in engineering systems

=) Decision support system.

3) Hidden Markov model (HMM):-

A hidden markov model (HMM) is a statistical model for system that evolve over time but where underlying states are hidden. Instead, we observe outcome generate probabilistic from hidden states.

components:-

a) Hidden states (S)

b) Observation (O)

c) Transition probabilities (A)

d) Emission probabilities (B)

e) Initial state distribution (π)

key problems solved by HMM;

=) Evaluation : compute probability of observation sequence given model.

=) Decoding : Determine most likely sequence of hidden states (viterbi algorithm)

=) Learning : Estimate parameters.

Application:-

=) Speech recognition

=) Natural language processing

=) Bio Informatics...

4) Bayesian Inference:-

Bayesian inference is a method of statical inference where probability is used to represent uncertainity about parameter.. unlike frequentist method, bayesian methods update beliefs based on new evidence

$$P(H|D) = \frac{P(D|H) \cdot P(H)}{P(D)}$$

where,

H -) hypothesis

D -) observed data.

advantages:-

=> Incoporates prior knowledge

=> Naturally handles uncertainty

=> produces full probability distribution.

Problem based on Naive bayes:-

A spam filter uses naive bayes algorithm consider the word "offer" appear in email.

P(spam) = 0.4

P(not spam) = 0.6

P(offer | spam) = 0.8

P(offer | not spam) = 0.2

If email contains word "offer", classify it as spam or not spam.

Solution:-

$$P(\text{spam} / \text{offer}) = \frac{P(\text{offer} / \text{spam})\, P(\text{spam})}{P(\text{offer})}$$

$$P(\text{offer}) = P(\text{offer} \mid \text{spam})\,P(\text{spam}) + P(\text{offer} \mid \text{not spam})\,P(\text{not spam})$$

$$= (0.8)(0.4) + (0.2)(0.6)$$

$$= 0.32 + 0.12$$

$$= 0.44$$

Now,

$$P(\text{spam} \mid \text{offer}) = \frac{(0.8 \times 0.4)}{0.44} = 0.727$$

$$P(\text{not spam} \mid \text{offer}) = \frac{0.2 \times 0.6}{0.44} = 0.273$$

∴ Here the email is classified as "spam".