Name : Tanya Verma

Roll No. : A023119820023

Class : 5AI-1

**Assignment**

Module - 1

**Fundamentals of Machine Learning (CSE313)**

Date :

---

$\boxed{1.}$

<u>Bayes' Theorem</u> :

In probability theory and statistics, Bayes' Theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event. It is a mathematical formula for determining conditional probability.

$$\underbrace{P(h|D)}_{\text{Posterior Probability}} = \frac{\overbrace{P(D|h)}^{\text{Likelihood}}\;\overbrace{P(h)}^{\text{Prior Probability}}}{\underbrace{P(D)}_{\text{Marginal Probability}}} \rightarrow \text{Likelihood Ratio}$$

, where

$P(h)$ = Probability of h occuring

$P(D)$ = Probability of D occuring

$P(h|D)$ = Probability of h given D

$P(D|h)$ = Probability of D given h

## Maximum A Posteriori (MAP) Estimate :

MAP Estimate is an estimate of an unknown quantity, that equals the mode of posterior distribution. The MAP can be used to obtain a point estimate of an unobserved quantity on the basis of empirical data.

$$\theta_{MAP} = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{m} \{\log p(x^{(i)} | \theta)\} + \log p(\theta)$$

$\boxed{2.}$

According to Bayes rule, the posterior can be decomposed into the product of the likelihood and prior.

$$\theta_{MAP} = \underset{\theta}{\operatorname{argmax}}\ p(\theta | X)$$

$$= \underset{\theta}{\operatorname{argmax}}\ p(X|\theta)\, p(\theta) \quad \text{(by applying bayes rule)}$$

$$= \underset{\theta}{\operatorname{argmax}}\ \log p(X|\theta) + \log p(\theta) \quad \text{(by applying log}$$

$$= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{m} \{\log p(x^{(i)} | \theta\} + \log p(\theta)$$

3.

## Naïve Bayes Classifier :

Naïve Bayes classifier algorithm is a supervised learning algorithm based on Bayes' Theorem which is used for solving classification problems.

It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.

It assumes that the occurence of a certain feature is independent of the occurence of other features.

$$P(X,Y) = P(X \wedge Y) = P(X \cap Y) = P(X \text{ and } Y) = \text{Joint Probability}$$

$$= P(X)P(Y)$$
$$= P(X|Y)\,P(Y)$$
$$= P(Y|X)\,P(X)$$

$$P(X,Y,Z) = P(X|Y,Z)\,P(Y|Z)\,P(Z)$$

4.

Let

TO = Event that Team 0 wins

T1 = Event that Team 1 win

T1F = Event that Team 1 hosts the game

So, we need to calc.

$P(T0 \mid T1H) = ?$

$P(T1 \mid T1H) = ?$ and compare these two.

Now,

$$P(T0 \mid T1H) + P(T1 \mid T1H) = 1$$

So, $$\frac{P(T1H \mid T0)P(T0)}{P(T1H)} + \frac{P(T1H \mid T1)P(T1)}{P(T1H)} = 1$$

$$\Rightarrow P(T1H) = P(T1H \mid T0)P(T0) + P(T1H \mid T1)P(T1)$$

$$P(T1H) = 0.30 \times 0.95 + 0.75 \times 0.05$$

$$P(T1H) = 0.3225$$

Now,

$$P(T0 \mid T1H) = \frac{P(T1H \mid T0)P(T0)}{P(T1H)} = \frac{0.30 \times 0.95}{0.3225}$$

$$P(T0 \mid T1H) = 0.88372$$

$$P(T1 \mid T1H) = \frac{P(T1H \mid T1)P(T1)}{P(T1H)} = \frac{0.75 \times 0.05}{0.3225}$$

$$P(T1 \mid T1H) = 0.11628$$

As $P(T0 \mid T1H) > P(T1 \mid T1H)$

=> Team 0 has a higher chance of winning the game if Team 1 hosts.

5.

Machine learning is a branch of Artificial Intelligence which focuses on studying and developing systems and algorithms that allows computer to recognize patterns & learn from data or the past to optimize a performance criterion and predict future outcomes.

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P.

The steps in designing learning system are:

1.) Choosing the training experience
2.) Choosing the target function
3.) Choosing ~~function approximation algorithm.~~ representation of target function

4) Choosing a function approximation algorithm
5) Find an evaluation procedure to test learned function

6.

Hypothesis space :

'h' is described by a conjunction of constraints on the attribute, the constraints may General hypothesis "?" (any value is acceptable), Specific hypothesis "ϕ" (a specific value or no value is accepted)

Instance space :

It is a subset of all possible examples or instances.

Version space :

The version space denotes $VS_{HD}$ (with respect to hypoth-esis space H and training example D) is the subset of hypothesis from H consistent with training example in D.

8.

Given,

$P(Y)$ = Probability of car stolen = $\frac{5}{10}$

$P(N)$ = Probability of car not stolen = $\frac{5}{10}$

| Color/Stolen | Y | N |
|---|---|---|
| Red | 3/5 | 2/5 |
| Yellow | 2/5 | 3/5 |

| Type / stolen | Y | N |
|---|---|---|
| Sports | 4/6 | 2/6 |
| SUV | 1/4 | 3/4 |

| Origin/ stolen | Y | N |
|---|---|---|
| Domestic | 2/5 | 3/5 |
| Imported | 3/5 | 2/5 |

New Instance = { Color = Red, Type = SUV, Origin = Domestic }

$P(Y \mid \text{New Instance}) = P(Y)\, P(\text{Red} \mid Y)\, P(\text{SUV} \mid Y)\, P(\text{Domestic} \mid Y)$

$= \frac{5}{10} \times \frac{3}{5} \times \frac{1}{4} \times \frac{2}{5}$

$= 0.03$

$P(N \mid \text{New Instance}) = P(N)\ \cancel{P(\ )}\ P(\text{Red} \mid N)\ P(\text{SUV} \mid N)\ P(\text{Domestic} \mid N)$

$= \frac{5}{10} \times \frac{2}{5} \times \frac{3}{4} \times \frac{3}{5}$

$= 0.09$

as $P(N \mid \text{New Instance}) > P(Y \mid \text{New Instance})$

The probability of vehicle with new instances ~~is~~ not getting stolen is higher than probability of vehicle with new instances getting stolen.