16/8/2022

B.Tech (AI)

CSE 313

HITESH

A023119820027

Fundamentals of Machine Learning

ASSIGNMENT

① Bayes Theorem is also known as the Bayes Rule or Bayes Law. It is a method to determine the probability of an event based on the occurrences of prior events.
It is used to calculate conditional probability.
Bayes theorem states that the conditional probability of an event A, given the occurrence of another event B, is equal to the product of the likelihood of B, given A and the probability of A. It is given as

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Where, $P(A)$ = how likely B happens (Marginalization)

$P(A/B)$ = how likely A happens given that B has happened (Posterior)

$P(B/A)$ = how likely B happens given that A has happened (Likelihood)

Maximum A Posteriori (MAP) hypothesis is the most probable hypothesis $h_{MAP} \in H$ (Candidate hypothesis) given the observed data D.

The posterior distribution, $f_{X|Y}(x|y)$ (or $P_{X|Y}(x|y)$), contains all the knowledge about the unknown quantity X. Therefore, we can use the posterior distribution to find point or interval estimates of X. One way to obtain a point estimate is to choose the value of x that maximizes the posterior PDF (or PMF). This is called Maximum a posteriori (MAP) estimation.

The MAP estimate of the random variable X, given that we have observed $Y = y$, is given by the value of x that maximizes

$f_{X|Y}(x|y)$ if X is a continuous random variable

$P_{X|Y}(x|y)$ if X is a discrete random variable

The MAP estimate is shown by $\hat{x}_{MAP}$.

(2) We can determine the MAP hypothesis by using Bayes theorem to calculate the posterior probability of each candidate hypothesis.

$$h_{MAP} = \underset{h \in H}{\operatorname{argmax}}\ P(h \mid D)$$

$$= \underset{h \in H}{\operatorname{argmax}}\ \frac{P(D \mid h)\, P(h)}{P(D)}$$

$$= \underset{h \in H}{\operatorname{argmax}}\ P(D \mid h)\, P(h)$$

Where $h \Rightarrow$ most probable hypothesis

$H \Rightarrow$ Candidate hypothesis

$D \Rightarrow$ Observed data

Note: $P(D)$ does not depend on the value of h. Thus, we can equivalently find the value of h that maxmizes

$$P(D \mid h)\, P(h)$$

This simplifies finding the MAP hypothesis because to find $P(D)$, the law of total probability is to be used which involves integration or summation.

(3) Naïve Bayes Classifier is one of the most simplest but powerful algorithms for classification based on Bayes Theorem with an assumption of independence among predictors.

It assumes that the presence of a feature in a class is unrelated to any other feature.

Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that a particular fruit is an apple or an orange, and that is why it is known as "Naïve".

Applications of Naïve Bayes Algorithm –

1) Face Recognition
2) Mail Classification
3) Handwriting Analysis
4) Salary Prediction

Naïve Bayes Classifier Formula –

$$\boxed{V_{NB} = \underset{v_j \in V}{\operatorname{argmax}} \, P(v_j) \prod_i P(a_i \mid v_j)}$$

(4) Probability that Team 0 wins $P(Y_0) = 0.95$

Probability that Team 1 wins $P(Y_1) = 1 - P(Y_0)$
$= 0.05$

Probability that team 1 hosted the match it had won,
$P(X_1 | Y_1) = 0.75$

Probability that team 1 hosted the match that is won by team 0, $P(X_1 | Y_0) = 0.30$

The Conditional Probability $P(Y_1 | X_1)$ that Team 1 wins the next match it hosts, can be calculated by using Bayes Theorem,

$$P(Y_1 | X_1) = \frac{P(X_1 | Y_1)\, P(Y_1)}{P(X_1)}$$

$$= \frac{P(X_1 | Y_1)\, P(Y_1)}{P(X_1 | Y_1) P(Y_1) + P(X_1 | Y_0) P(Y_0)}$$

$$= \frac{0.75 \times 0.05}{(0.75 \times 0.05) + (0.30 * 0.95)}$$

$P(Y_1 | X_1) = 0.1162$

$P(Y_0 | X_1) = 1 - P(Y_1 | X_1)$
$= 0.8838$

There are 2 events - win & host
$Y \Rightarrow$ win
$X \Rightarrow$ host
$0 \Rightarrow$ team 0
$1 \Rightarrow$ team 1

$\because P(Y_1 | X_1) < P(Y_0 | X_1)$

$\therefore$ Team 0 has a better probability of winning than Team 1.

⑤ Machine Learning (ML) is a branch of Artificial Intelligence, concerned with the design and development of algorithms that allow computers to evolve behaviours based on empirical data.

As intelligence requires knowledge, it is necessary for the computers to acquire knowledge.

The steps in designing a learning system are as follows -

1) Training Set - Choosing the training experience (training set), i.e., training inputs - outputs and how to represent it.

2) Target Function - Choose how to represent the target function to learn the best move.

3) Learning Algorithm - Choose the learning algorithm to infer the target from experience (for achieving more accuracy).

4) Evaluation Procedure - Find an evaluation procedure and matrices to test learned function.

5) Final Design - The final design is created at last when system goes from a no. of examples, failures & success.

⑥ Hypothesis Space 'h' is defined by a conjunction of constraints on the attribute, the constraints may be General hypothesis "?" (any value is acceptable), or Specific hypothesis "ϕ" (a specific value or no value is accepted).

Instance Space is a subset of all possible examples or instance.

Version Space denotes VS_{HD} (with respect to hypothesis space H and training example D) is the subset of hypothesis from H consistent with training example in D.

---

⑧ Applying Naïve Bayes Classifier,

$$V_{NB} = \underset{v_j \in \{yes, no\}}{\operatorname{argmax}} P(v_j) \prod_i P(a_i | v_j)$$

(Attributes, Values) ⇒ (Color | Red, yellow)
(Type | Sports, SUV)
(Origin | Domestic, Imported)

Target Class

	Color	Yes	No
Values	Red	3	2
	Yellow	2	3

$P(\text{color} = \text{Red} \mid \text{Stolen} = \text{Yes}) = \frac{3}{5} = 0.6$

$P(\text{color} = \text{Red} \mid \text{Stolen} = \text{No}) = \frac{2}{5} = 0.4$

$P(\text{color} = \text{Yellow} \mid \text{Stolen} = \text{Yes}) = \frac{2}{5} = 0.4$

$P(\text{color} = \text{Yellow} \mid \text{Stolen} = \text{No}) = \frac{3}{5} = 0.6$

Type	Yes	No
Sports	4	2
SUV	1	3

Target Class

Values

$P(\text{Type} = \text{Sports} \mid \text{Stolen} = \text{yes}) = \frac{4}{5} = 0.8$

$P(\text{Type} = \text{sports} \mid \text{Stolen} = \text{No}) = \frac{2}{5} = 0.4$

$P(\text{Type} = \text{SUV} \mid \text{Stolen} = \text{Yes}) = \frac{1}{5} = 0.2$

$P(\text{Type} = \text{SUV} \mid \text{Stolen} = \text{No}) = \frac{3}{5} = 0.6$

Origin	Yes	No
Domestic	2	3
Imported	3	2

Target Class

Values

$P(\text{Origin} = \text{Domestic} \mid \text{Stolen} = \text{Yes}) = \frac{2}{5} = 0.4$

$P(\text{Origin} = \text{Domestic} \mid \text{Stolen} = \text{No}) = \frac{3}{5} = 0.6$

$P(\text{Origin} = \text{Imported} \mid \text{Stolen} = \text{Yes}) = \frac{3}{5} = 0.6$

$P(\text{Origin} = \text{Imported} \mid \text{Stolen} = \text{No}) = \frac{2}{5} = 0.4$

Classify the new instance = (Red, SUV, Domestic)

For Stolen = Yes, P(Stolen = Yes | New Instance)

$\Rightarrow P(\text{Color} = \text{Red} \mid \text{Stolen} = \text{Yes}) P(\text{Type} = \text{SUV} \mid \text{Stolen} = \text{Yes}) P(\text{Origin} = \text{Domestic} \mid \text{Stolen} = \text{Yes}) \times P(\text{Yes})$

$\Rightarrow 0.6 \times 0.2 \times 0.4 \times 0.5$

$\Rightarrow 0.024$

For Stolen = No, P(Stolen = No | New Instance)

$\Rightarrow P(\text{color} = \text{Red} \mid \text{Stolen} = \text{No}) P(\text{Type} = \text{SUV} \mid \text{Stolen} = \text{No}) P(\text{Origin} = \text{Domestic} \mid \text{stolen} = \text{No}) P(\text{No})$

$\Rightarrow 0.4 \times 0.6 \times 0.6 \times 0.5$

$\Rightarrow 0.072$

$\because P(\text{Stolen} = \text{No} \mid \text{new instance}) > P(\text{Stolen} = \text{Yes} \mid \text{new instance})$

$\therefore$ The new instance would classified as Not Stolen.