NAIVE BAYESI

1. It is a classification based model.
2. Naive Bayes works on supervised data where, class variable is discrete in nature.
3. It is a probabilistic model.
4. It uses Bayes' theorem to make predictions.

PROBABILISTIC CONCEPTS I

Probability is the measure to capture uncertainty. It is used to find the likelihood of an event to occur. For example, the weather report might say "there is high probability of rain today". Probabilistic theory deals with such estimates and the rules which they should obey. Probability theory works on:

1. **Experiments:** The subject.
   For example, "coin flip", "cards", "rolling a dice".

2. **Events/ Outcome:** The result of experiment. For example, "Getting a head in coin flip", "Getting 6 in rolling a dice". The event is generally represented by α.

3. **Event Space:** The set of possible events in the experiment.

   For example, "in coin flip the event space is 2". In experiment of "rolling a dice, the event space is 6". It is generally denoted by Ω.

   The $\Omega = \{ \text{Head, Tail} \}$ on coin flip problem whereas, $\Omega = \{ 1,2,3,4,5,6 \}$ on rolling a dice problem.

PROBABILISTIC CONCEPTS II

Computing Probability of an Event

The probability of an event α in an experiment is computed using Equation 19. Where P stands for probability.

$$P(\alpha) = \frac{\textit{Number of favourable outcome}}{\textit{Total number of outcomes}} \qquad (19)$$

Examples

1. **Let Ω = {Head, Tail}.**
   Here, P (Head) = ½ and P (Tail) =½.

2. **Let Ω = { male, male, male, female, male, female, male, female, female, male}**
   Ω represents data on gender of a class with 10 students.

   Here, P (male) = 6/10 and P (female) = 4/10.

PROBABILISTIC CONCEPTS III

Conditional Probability

Conditional probability is a measure of the probability of an event α given that another event say β has occurred. The conditional probability is computed using Equation 20.

$$P(\alpha|\beta) = \frac{P(\beta \cup \alpha)}{P(\beta)} \qquad (20)$$

Examples

The probability that it is Friday and that it is party 0.03. The probability that it is Friday is 0.14 (1/7). What is the probability that it is party given it is Friday ?

$$P(Party \mid Friday) = \frac{P(Party \cup Friday)}{P(Friday)} = \frac{0.03}{0.14} = 0.214 \approx 21\% \qquad (21)$$

$\alpha = Party$ $\beta = Friday$

- Bayes' rule describes the probability of an event, based on prior knowledge of conditions that might be related to the event.
- Bayes' rule is more useful when conditional probability is estimated directly from data.

PROBABILISTIC CONCEPTS IV

Bayes' Theorem

Bayes' theorem is used to determine conditional probability. The theorem provides a way to revise existing predictions or theories given new or additional evidence. It is common to think of Bayes' rule in terms of updating our belief in α in the light of new evidence β. Bayes' theorem is expressed in Equation 22.

$$P(\alpha|\beta) = \frac{P(\beta|\alpha) \text{ X } P(\alpha)}{P(\beta)} \tag{22}$$

The power of Bayes' rule is computing P $(\alpha|\beta)$. Which in many situations is difficult to compute directly. However, we may directly compute P $(\beta|\alpha)$. Bayes' rule enables us to compute P $(\alpha|\beta)$ in terms of P $(\beta|\alpha)$.

PROBABILISTIC CONCEPTS V

Example

Suppose that we are interested in diagnosing cancer in patients who visit a clinic. Let,

1. α represents the event person has cancer.
2. β represents the event person is a smoker.
3. On the bases of past data, let we have prior probability of events, P (α) = 0.1 and, P (β) = 0.5.
4. We are interested in knowing probability of person suffering from cancer given he smokes, *i.e.*, P ($\alpha \mid \beta$).
5. It is difficult to solve this query directly. However, we are likely to know P ($\beta \mid \alpha$) from records specifying the proportion of smokers among those diagnosed. Suppose P ($\beta \mid \alpha$) = 0.8.
6. Compute P ($\alpha \mid \beta$) using Bayes' rule,

$$P(\alpha|\beta) = P(\text{Cancer} \mid \text{Smoker}) = \frac{P(\text{Smoker} \mid \text{Cancer}) \times P(\text{Smoker})}{P(\text{Cancer})} = \frac{0.8 \text{ X } 0.1}{0.5} = 0.16 \qquad (23)$$

PROBABILISTIC CONCEPTS VI

Bayes' Theorem

1. Computing P ($\alpha | \beta$) is fundamental in machine learning to make predictions based on probabilistic theory.

2. Using Bayes' rule, we can find interesting prediction such as,

   P(Email = spam | X) =?

   P(Weather = rainy | X) =?

3. Bayes' rule is foundation of Naive Bayes classifier.

NAIVE BAYES FOR CLASSIFICATION I

1. Consider we are given supervised data set.
2. The supervised data set contains set of indicator variables and a class variable. We represent set of indicator variables by X and, class variable by C.
3. It is mandatory that C is categorical in nature but X can be mixture of numerical and categorical variables.
4. The goal of Naive Bayes classification is to estimate/compute: P(C|X).
   More formally, the objective is find:
   $P (C \mid x_1, x_2, x_3, \ldots, x_n)$, where $X = (x_1, x_2, x_3, \ldots, x_n)$
5. Where $P (C \mid x_1, x_2, x_3, \ldots, x_n)$ is computed using the data set given and Bayes' rule.

α <-> C
β <-> X
$P(\alpha \mid \beta)$
$P(C \mid X)$
X=(x1,x2,x3……..xn)