AI2002 – ARTIFICIAL INTELLIGENCE SPRING 2024 - LECTURE 31-33

Presented By: Mr. Sandesh Kumar
Following slides belong to Nitin Agarwal, Software Consultant

Agenda

- What is Conditional Probability ?
- What is Bayes Theorem?
- What is NAIVE BAYES CLASSIFIER?
- Types of Naive Bayes Algorithm.

Conditional Probability

1. In probability theory, conditional probability is a measure of the probability of an event given that another event has already occurred.

2. If the event of interest is A and the event B is assumed to have occurred, "the conditional probability of A given B", or "the probability of A under the condition B", is usually written as P(A|B), or sometimes PB(A).

Examples

Chances of cough

The probability that any given person has a cough on any given day maybe only 5%. But if we know or assume that the person has a cold, then they are much more likely to be coughing. The conditional probability of coughing given that person have a cold might be a much higher 75%.

Marbles in a Bag

2 blue and 3 red marbles are in a bag.

What are the chances of getting a blue marble?

???

Marbles in a Bag

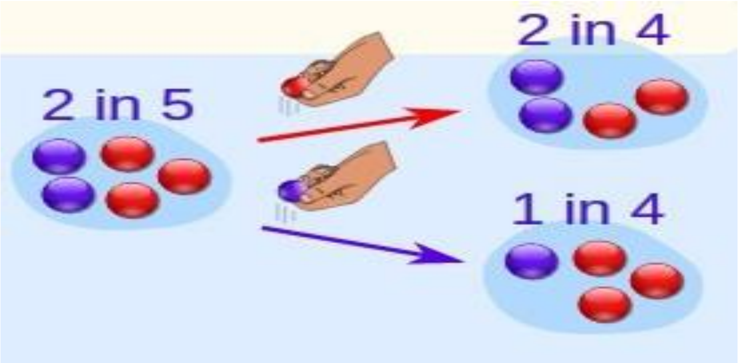
2 blue and 3 red marbles are in a bag.

What are the chances of getting a blue marble?

Answer: - The chance is 2 in 5

But after taking one out of these chances,

situation may change!



So the next time:

1. if we got a red marble before, then the chance of a blue marble next is 2 in 4
2. if we got a blue marble before, then the chance of a blue marble next is 1 in 4

Likewise:-

Drawing a second ace from a deck given we got the first ace

Finding the probability of having a disease given you were tested positive

Finding the probability of liking Harry Potter given we know the person likes fiction.

Bayes Theorem

1. In probability theory and statistics, Bayes' theorem (alternatively Bayes' law or Bayes' rule) describes the probability of an event, based on prior knowledge of conditions that might be related to the event.

2. For example, if cancer is related to age, then, using Bayes' theorem, a person's age can be used to more accurately to assess the probability that they have cancer, compared to the assessment of the probability of cancer made without knowledge of the person's age.

The Formula for Bayes' theorem

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

where

1. P(H) is the probability of hypothesis H being true. This is known as the prior probability.
2. P(E) is the probability of the evidence(regardless of the hypothesis).
3. P(E|H) is the probability of the evidence given that hypothesis is true.
4. P(H|E) is the probability of the hypothesis given that the evidence is there.

Examples

Suppose there are three bowls B1, B2, B3 and bowl B1 has 2 red and 4 blue coins; bowl B2 has 1 red and 2 blue coins; bowl B3 contains 5 red and 4 blue coins.Suppose the probabilities for selecting the bowls is not the same but are:-

- P(B1) = 1/3
- P(B2) = 1/6
- P(B3) = 1/2

Now, let us compute, assuming that a red coin was drawn what will be the probability that it came from bowl B1.

In mathematics teams, we need to find out **P(B1|R) = ???**

<u>And according to Bayes' theorem</u>

P(B1|R) = P(R|B1) * P(B1) / P(R)

For that first, we need to calculate some probabilities which are:-

- Probability to select a red coin i.e **P(R)**
- Probability to select the bowl 1 (B1) i.e **P(B1)** which is already given 1/3
- Probability to select a red coin from B1 i.e **P(R|B1)**

P(R)

$= P(B1 \cap R) + (B2 \cap R) + P(B3 \cap R)$

where

- $P(B1 \cap R)$ is probability to select bowl 1 and red coin
- $P(B2 \cap R)$ is probability to select bowl 2 and red coin
- $P(B3 \cap R)$ is probability to select bowl 3 and red coin

= P(selecting B1) * P(Number of Red coins / total number of coins in B1) + P(selecting B2) * P(Number of Red coins in B2 / total number of coins in B2) + P(selecting B3) * P(Number of Red coins in B3/ total number of coins in B3)

= 1/3 * 2/6 + 1/6 * 1/3 + 1/2* 5/9

= 4/9

So P(R) = 4/9

P(R|B1)

The probability of selecting a red coin given that it will be drawn from B1 is **2/6**

P(B1) was given i.e 1/3.

By putting all the values in formula:

P(B1|R) = (2/6 *1/3) / 4/9 = 2/8 = 0.25

So we can say that if a red coin was drawn that it will be 25% chances that it was drawn from bowl 1 i.e B1.

NAIVE BAYES CLASSIFIER

- Naive Bayes is a kind of classifier which uses the Bayes Theorem.

- It predicts membership probabilities for each class such as the probability that given record or data point belongs to a particular class.

- The class with the highest probability is considered as the most likely class. This is also known as **Maximum A Posteriori (MAP).**

$MAP(H)$

$= max(P(H|E))$

$= max((P(E|H)*P(H))/P(E))$

$= max(P(E|H)*P(H))$

P(E) is evidence probability, and it is used to normalize the result. It remains same so, removing it won’t affect.

Assumption

Naive Bayes classifier assumes that all the features are unrelated to each other. Presence or absence of a feature does not influence the presence or absence of any other feature.

“A fruit may be considered to be an apple if it is red, round, and about 4" in diameter. Even if these features depend on each other or upon the existence of the other features, a naive Bayes classifier considers all of these properties to independently contribute to the probability that this fruit is an apple.”

In real datasets, we test a hypothesis given multiple evidence(feature). So, calculations become complicated. To simplify the work, the feature independence approach is used to 'uncouple' multiple evidence and treat each as an independent one.

P(H|Multiple Evidences) = P(E1| H)* P(E2|H) ……*P(En|H) * P(H) / P(Multiple Evidences)

For understanding a theoretical concept, the best procedure is to try it on an example.

Let’s consider a training dataset with 1500 records and 3 classes. We presume that there are no missing values in our data. We have

We have 3 classes associated with Animal Types:

- Parrot,
- Dog,
- Fish.

The Predictor features set consists of 4 features as

- Swim
- Wings
- Green Color
- Dangerous Teeth.

Swim,Wings,Green Color, Dangerous Teeth. All the features are categorical variables with either of the 2 values: **T(True) or F(False).**

Swim	Wings	Green Color	Dangerous Teeth	Animal Type
50	500/500	400/500	0	Parrot
450/500	0	0	500/500	Dog
500/500	0	100/500	50/500	Fish

The above table shows a frequency table of our data. In our training data:

- Parrots have 50(10%) value for Swim, i.e., 10% parrot can swim according to our data, 500 out of 500(100%) parrots have wings, 400 out of 500(80%) parrots are Green and 0(0%) parrots have Dangerous Teeth.
- Classes with Animal type Dogs shows that 450 out of 500(90%) can swim, 0(0%) dogs have wings, 0(0%) dogs are of Green color and 500 out of 500(100%) dogs have Dangerous Teeth.
- Classes with Animal type Fishes shows that 500 out of 500(100%) can swim, 0(0%) fishes have wings, 100(20%) fishes are of Green color and 50 out of 500(10%) Fishes have Dangerous Teeth.

Now, it’s time to work on predict classes using the Naive Bayes model.

We have taken 2 records that have values in their feature set, but the target variable needs to predicted.

	Swim	Wings	Green	Teeth
1.	True	False	True	False
2.	True	False	True	True

- We have to predict animal type using the feature values. We have to predict whether the animal is a Dog, a Parrot or a Fish

P(H|Multiple Evidences) = P(E1| H)* P(E2|H) ……*P(En|H) * P(H) / P(Multiple Evidences)

Let’s consider the first record.

The Evidence here is Swim & Green. The Hypothesis can be an animal type to be Dog, Parrot, Fish.

For Hypothesis testing for the animal to be a Dog:

P(Dog | Swim, Green) = P(Swim|Dog) * P(Green|Dog) * P(Dog) / P(Swim, Green)

= 0.9 * 0 * 0.333 / P(Swim, Green)

= 0

For Hypothesis testing for the animal to be a Parrot:

P(Parrot| Swim, Green) = P(Swim|Parrot) * P(Green|Parrot) * P(Parrot) / P(Swim, Green)

= 0.1 * 0.80 * 0.333 / P(Swim, Green)

= 0.0264/ P(Swim, Green)

For Hypothesis testing for the animal to be a Fish:

P(Fish| Swim, Green) = P(Swim|Fish) * P(Green|Fish) * P(Fish) / P(Swim, Green)

= 1 * 0.2 * 0.333 / P(Swim, Green)

= 0.0666/ P(Swim, Green)

The denominator of all the above calculations is same i.e, P(Swim, Green).
The value of P(Fish| Swim, Green) is greater that P(Parrot| Swim, Green).

Using Naive Bayes, we can predict that the class of this record is Fish.

Let’s consider the second record.

The Evidence here is Swim, Green & Teeth. The Hypothesis can be an animal type to be Dog, Parrot, Fish.

For Hypothesis testing for the animal to be a Dog:

P(Dog | Swim, Green, Teeth) = P(Swim|Dog) * P(Green|Dog) * P(Teeth|Dog) * P(Dog) / P(Swim, Green, Teeth)

= 0.9 * 0 * 1 * 0.333 / P(Swim, Green, Teeth) = 0

For Hypothesis testing for the animal to be a Parrot:

P(Parrot| Swim, Green, Teeth) = P(Swim|Parrot) * P(Green|Parrot)* P(Teeth| Parrot) * P(Parrot) / P(Swim, Green, Teeth)

= 0.1 * 0.80 * 0 *0.333 / P(Swim, Green, Teeth)

= 0

For Hypothesis testing for the animal to be a Fish:

P(Fish|Swim, Green, Teeth) = P(Swim|Fish) * P(Green|Fish) * P(Teeth|Fish) *P(Fish) / P(Swim, Green, Teeth)

= 1 * 0.2 * 0.1 * 0.333 / P(Swim, Green, Teeth)

= 0.00666 / P(Swim, Green, Teeth)

The value of P(Fish| Swim, Green, Teeth) is the only positive value greater than 0. Using Naive Bayes, we can predict that the class of this record is **Fish.**

Types of Naive Bayes Algorithm

Gaussian: It is used in classification and it assumes that features follow a normal distribution.

Bernoulli: The binomial model is useful if your feature vectors are binary (i.e. zeros and ones). One application would be text classification with ‘bag of words’ model where the 1s & 0s are “word occurs in the document” and “word does not occur in the document” respectively.

MultiNomial Naive Bayes is preferred to use on data that is multinomially distributed.

For example, let’s say, we have a text classification problem. Here we can consider bernoulli trials which is one step further and instead of “word occurring in the document”, we have “count how often word occurs in the document”, you can think of it as “number of times outcome number x_i is observed over the n trials”.