Naïve Bayes

What we will Learn..

- Naïve Bayes
- Classification Approach of Naïve Bayes
- How does Naïve Bayes Classifier work?
- Probabilistic Measures used in Naïve Bayes
  - Prior Probability
  - Conditional Probability
  - Posterior Probability
- Classifying a new/unseen datapoint using Naïve Bayes
- Advantages and Disadvantages of Naïve Bayes
- Applications of Naïve Bayes

Naïve Bayes

Naïve Bayes is a probabilistic technique that is used for Classification task(s) which is based on the Bayes Theorem.

Assumption(s) in Naïve Bayes Classification Approach

Outlook	Temperature	Humidity	*PLAY GOLF*
Sunny	Hot	High	No
Overcast	Hot	Normal	No
Overcast	Cool	High	Yes
Rain	Mild	Normal	No
Sunny	Cool	Normal	Yes
Rain	Cool	Normal	No

Let’s consider the First row from the Dataset above:

- Temperature “Hot” doesn’t imply that Humidity is “High”. Likewise, Outlook “Overcast” doesn’t imply the Temperature is “Hot”
- Every predictor has equal effect on the “Outcome”. That means, Humidity “High” can not solely decide whether Golf will be played or, not.

Assumption(s) in Naïve Bayes Classification Approach Continued…

Outlook	Temperature	Humidity	*PLAY GOLF*
Sunny	Hot	High	No
Overcast	Hot	Normal	No
Overcast	Cool	High	Yes
Rain	Mild	Normal	No
Sunny	Cool	Normal	Yes
Rain	Cool	Normal	No

The assumptions are, the Predictors are **Independent**, they don’t affect other predictors and the predictors **equally** contribute to the outcome.

Hence, it is called “**Naïve**”

What does Naïve Bayes Classifier do?

Outlook	Temperature	Humidity	*PLAY GOLF*
Sunny	Hot	High	No
Overcast	Hot	Normal	No
Overcast	Cool	High	Yes
Rain	Mild	Normal	No
Sunny	Cool	Normal	Yes
Rain	Cool	Normal	No

Overcast	**Hot**	**High**	**???**

Now, can the "Naïve Bayes Classifier predict whether Golf will be Played "Yes" or not Played "No" for the new unseen predictor given here?

How does Naïve Bayes Classifier work?

Naïve Bayes Classifier combines the ***Prior Probability*** and ***Conditional Probability*** which can then be used for each possible Classification.

Having done so, we then choose the Classification with the Largest Probability.

Prior Probability in Naïve Bayes

Prior Probability is the probability of an event that is measured without dependency on any other values (such as, predictors).

Note: Prior Probability is estimated before any other information in the Dataset is considered.

Prior Probability in Naïve Bayes Continued…

Outlook	Temperature	Humidity	*PLAY GOLF*
Sunny	Hot	High	No
Overcast	Hot	Normal	Yes
Overcast	Cool	High	Yes
Rain	Mild	Normal	No
Sunny	Cool	Normal	Yes
Rain	Cool	Normal	No
Overcast	Cool	Normal	Yes

The probability of Golf being played is calculated using the frequency of “Yes” based on the total occurrences is the Prior Probability.

Prior Probability of “Yes”= 4/7 = 0.57
Prior Probability of “No”= 3/7 = 0.43

Conditional Probability in Naïve Bayes

Conditional Probability is a Measure of the probability of an Event given that, another event has already occurred.

$$P= (A \mid B)$$

Means, “A” will occur given that “B” has already occurred.

Note: Conditional Probability of an event is estimated based on when it is known that it has dependency/dependencies on any condition.

Conditional Probability in Naïve Bayes Continued…

Outlook	Temperature	Humidity	*PLAY GOLF*
Sunny	Hot	High	No
Overcast	Hot	Normal	Yes
Overcast	Cool	High	Yes
Rain	Mild	Normal	No
Sunny	Cool	Normal	Yes
Rain	Cool	Normal	No
Overcast	Cool	Normal	Yes

The probability of the Condition Outlook being “overcast” when Golf is Played can be estimated from the frequency of these mentioned instances in the dataset. For example,

Probability of (Outlook= Overcast | PLAY GOLF= Yes) = 3/4= 0.75

Similarly,

Probability of (Temperature= Hot | PLAY GOLF = Yes) = 1/4 =0.25

Probability of (Humidity = High | PLAY GOLF = Yes) = 1/4 = 0.25

Same estimation can be done for Golf is not played, or, “No” as well.

Posterior Probability: Combination of Prior and Conditional Probability

In Naïve Bayes, Posterior Probability is a revised probability of an event that occurs with the dependency of other values (for instance, predictors) along with the initially estimated prior probability.

Let’s see the next slide!!

Posterior Probability: Combination of Prior and Conditional Probability Continued…

For a k mutually exclusive and exhaustive Classifications $c_1, c_2, ...c_k$ the prior probabilities are, $P(c_1), P(c_2), P(c_k)$ respectively, n attributes $a_1, a_2, \ldots\ldots a_n$ which for a given instance have values $v_1, v_2, \ldots. v_n$ respectively. So, the Posterior Probability of class c_i for an instance can be shown as:

$$P(c_i) \times P(a_1 = v_1 \text{ and } a_2 = v_2 \ldots \text{ and } a_n = v_n \mid c_i)$$

Now, the assumption of the attributes being independent, the value of the expression can be:

$$\boldsymbol{P(c_i) \times P(a_1 = v_1 \mid c_i) \times P(a_2 = v_2 \mid c_i) \times \ldots \times P(a_n = v_n \mid c_i)}$$

Posterior Probability in Naïve Bayes Example

Outlook	Temperature	Humidity	*PLAY GOLF*
Sunny	Hot	High	No
Overcast	Hot	Normal	Yes
Overcast	Cool	High	Yes
Rain	Mild	Normal	No
Sunny	Cool	Normal	Yes
Rain	Cool	Normal	No
Overcast	Cool	Normal	Yes

Now, Let's determine the Posterior Probability for the mutually exclusive and exhaustive Class values "Yes" and "No" to predict the Class Value for the following unseen/ new Datapoint:

Overcast	Hot	High	???

Posterior Probability in Naïve Bayes Example Continued…
Play Golf "Yes"

Outlook	Temperature	Humidity	*PLAY GOLF*
Sunny	Hot	High	No
Overcast	Hot	Normal	Yes
Overcast	Cool	High	Yes
Rain	Mild	Normal	No
Sunny	Cool	Normal	Yes
Rain	Cool	Normal	No
Overcast	Cool	Normal	Yes

Overcast	Hot	High	???

Class Value: "PLAY GOLF" as "Yes":

Prior Probability:

4/7 = 0.57

Conditional Probabilities:

(Outlook= Overcast | PLAY GOLF= Yes) = 3/4= 0.75

(Temperature= Hot | PLAY GOLF = Yes) = 1/4 =0.25

(Humidity = High | PLAY GOLF = Yes) = 1/4 = 0.25

Posterior Probability:

0.57X 0.75 X 0.25 X 0.25 = **0.027**

Posterior Probability in Naïve Bayes Example Continued…
Play Golf “No”

Outlook	Temperature	Humidity	*PLAY GOLF*
Sunny	Hot	High	No
Overcast	Hot	Normal	Yes
Overcast	Cool	High	Yes
Rain	Mild	Normal	No
Sunny	Cool	Normal	Yes
Rain	Cool	Normal	No
Overcast	Cool	Normal	Yes

Overcast	Hot	High	???

Class Value: “PLAY GOLF” as “No”:

Prior Probability:

3/7 = 0.43

Conditional Probabilities:

(Outlook= Overcast | PLAY GOLF= No) = 0/3= 0

(Temperature= Hot | PLAY GOLF = No) = 1/3 =0.33

(Humidity = High | PLAY GOLF = No) = 1/3 = 0.33

Posterior Probability:

0.43 X 0 X 0.33 X 0.33 = 0

Comparing Posterior Probability for both "Yes" and "No"

Outlook	Temperature	Humidity	*PLAY GOLF*
Sunny	Hot	High	No
Overcast	Hot	Normal	Yes
Overcast	Cool	High	Yes
Rain	Mild	Normal	No
Sunny	Cool	Normal	Yes
Rain	Cool	Normal	No
Overcast	Cool	Normal	Yes

Posterior Probability for "Yes": 0.027

Posterior Probability for "No": 0

"Yes" has Larger Posterior Probability value that implies the Classification for the new/unseen Datapoint will be **"Yes"**, golf will be played.

Overcast	Hot	High	**Yes**

Advantages of Naïve Bayes

- Naïve Bayes is easy to implement
- Very Fast and appropriate where speedy training is important.
- Solve multiclass problem
- If the “Assumption” about the independent variables hold true, then Naïve Bayes works better than many other models.

Disadvantages of Naïve Bayes

- If the Test Data has a value that has never been observed hence, trained from the training set then it may cause Zero “0” probability
- The assumption of “mutually independent” attribute may not be always possible to find in real life

Applications of Naïve Bayes Algorithm

Applications where Naïve Bayes can be applied:

- Sentiment Analysis
- Spam identification
- Face Recognition
- Recommendation System
- Weather prediction
- News Classification
- Medical Diagnosis

and many more…


Do you have any
QUESTIONS?