1. Naive Bayes Algorithm

- Naive Bayes classifiers are a collection of classification algorithms based on Baye's Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.
- To start with, let us consider a dataset.
- Consider a fictional dataset that describes the weather conditions for playing a game of golf. Given the weather conditions, each tuple classifies the conditions as fit("Yes") or unfit("No") for plaing golf.

Assumption:

The fundamental Naive Bayes assumption is that each feature makes an:

- Independent contribution to the outcome.
- Equal contribution to the outcome. With relation to our dataset, this concept can be understood as:

We assume that no pair of features are dependent.

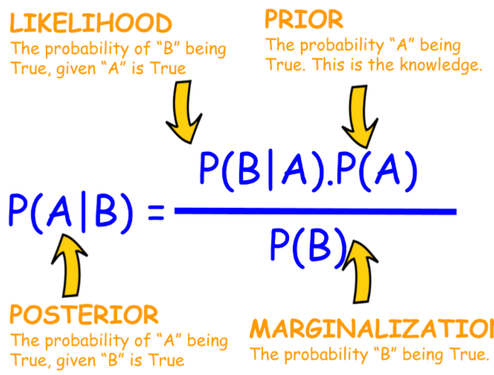
- For example, the temperature being 'Hot' has nothing to do with the humidity or the outlook being 'Rainy' has no effect on the winds. Hence, the features are assumed to be independent.

Each feature is given the same weight(or importance).

- For example, knowing only temperature and humidity alone can't predict the outcome accuratey. None of the attributes are irrelevant and assumed to be contributing equally to the outcome.
- Note: The assumptions made by Naive Bayes are not generally correct in real-world situations. In-fact, the independence assumption is never correct but often works well in practice.
- Now, before moving to the formula for Naive Bayes, it is important to know about Bayes' theorem.

Bayes' Theorem.

- Bayes' Theorem finds the probability of an event occurring given the probability of another event that has already occurred. Bayes' theorem is stated mathematically as the following equation:



- Where A and B are events and P(B) ? 0.
- Basically, we are trying to find probability of event A, given the event B is true. Event B is also termed as evidence.
- P(A) is the priori of A (the prior probability, i.e. Probability of event before evidence is seen). The evidence is an attribute value of an unknown instance(here, it is event B).
- P(A|B) is a posteriori probability of B, i.e. probability of event after evidence is seen.
- Now, with regards to our dataset, we can apply Bayes' theorem in following way:
- Just to clear, an example of a feature vector and corresponding class variable can be: (refer 1st row of dataset)
  - X = (Rainy, Hot, High, False)
  - y = No

Use the Naive Assumptions

- Now, its time to put a naive assumption to the Bayes' theorem, which is, **independence** among the features. So now, we split **evidence** into the independent parts.
- Now, if any two events A and B are independent, then,
  - P(A,B) = P(A)P(B)
- Please note that **P(y) is also called class probability** and **P(xi | y) is called conditional probability.**
- The different naive Bayes classifiers differ mainly by the assumptions they make regarding the distribution of data.
- P(xi | y).
- Let us try to apply the above formula manually on our weather dataset. For this, we need to do some precomputations on our dataset.
- We need to find **P(xi | yj)** for each **xi** in **X** and **yj** in **Y**. All these calculations have been demonstrated in the tables below:

Outlook

	Yes	No	P(yes)	P(no)
Sunny	2	3	2/9	3/5
Overcast	4	0	4/9	0/5
Rainy	3	2	3/9	2/5
Total	9	5	100%	100%

Temperature

	Yes	No	P(yes)	P(no)
Hot	2	2	2/9	2/5
Mild	4	2	4/9	2/5
Cool	3	1	3/9	1/5
Total	9	5	100%	100%

Humidity

	Yes	No	P(yes)	P(no)
High	3	4	3/9	4/5
Normal	6	1	6/9	1/5
Total	9	5	100%	100%

Wind

	Yes	No	P(yes)	P(no)
False	6	2	6/9	2/5
True	3	3	3/9	3/5
Total	9	5	100%	100%

Play		P(Yes)/P(No)
Yes	9	9/14
No	5	5/14
Total	14	100%

Principle of Naive Bayes Classifier:

A Naive Bayes classifier is a probabilistic machine learning model that's used for classification task. The crux of the classifier is based on the Bayes theorem.

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

Using Bayes theorem, we can find the probability of A happening, given that B has occurred. Here, B is the evidence and A is the hypothesis. The assumption made here is that the predictors/features are independent. That is presence of one particular feature does not affect the other. Hence it is called naive.

Types of Naive Bayes Classifier:

Multinomial Naive Bayes:

- This is mostly used for document classification problem, i.e whether a document belongs to the category of sports, politics, technology etc. The features/predictors used by the classifier are the frequency of the words present in the document.

Bernoulli Naive Bayes:

- This is similar to the multinomial naive bayes but the predictors are boolean variables. The parameters that we use to predict the class variable take up only values yes or no, for example if a word occurs in the text or not.

Gaussian Naive Bayes:

- When the predictors take up a continuous value and are not discrete, we assume that these values are sampled from a gaussian distribution.