Naive Bayes classification

- Naive Bayes classifiers are a collection of classification algorithms based on **Bayes' 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.

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.

Naive Bayes classification

	Outlook	Temperature	Humidity	Windy	Play Golf
0	Rainy	Hot	High	False	No
1	Rainy	Hot	High	True	No
2	Overcast	Hot	High	False	Yes
3	Sunny	Mild	High	False	Yes
4	Sunny	Cool	Normal	False	Yes
5	Sunny	Cool	Normal	True	No
6	Overcast	Cool	Normal	True	Yes
7	Rainy	Mild	High	False	No
8	Rainy	Cool	Normal	False	Yes
9	Sunny	Mild	Normal	False	Yes
10	Rainy	Mild	Normal	True	Yes
11	Overcast	Mild	High	True	Yes
12	Overcast	Hot	Normal	False	Yes
13	Sunny	Mild	High	True	No

Naive Bayes classification

- The dataset is divided into two parts, namely, **feature matrix** and the **response vector.**
- Feature matrix contains all the vectors(rows) of dataset in which each vector consists of the value of dependent features. In above dataset, features are ‘Outlook’, ‘Temperature’, ‘Humidity’ and ‘Windy’.
- Response vector contains the value of class variable(prediction or output) for each row of feature matrix. In above dataset, the class variable name is ‘Play golf’.

Naive Bayes classification

Assumption:

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

- independent
- equal

- 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.
- Secondly, each feature is given the same weight(or importance). For example, knowing only temperature and humidity alone can’t predict the outcome accurately. None of the attributes is irrelevant and assumed to be contributing equally to the outcome.

Naive Bayes classification

•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:

$$P(A|B) = P(B|A) \,.\, P(A) \;/\; P(B)$$

Now, with regards to our dataset, we can apply Bayes' theorem in following way:

$$P(ci|\Theta) = P(\Theta\, |ci).\, P(ci) \;/\; P(\Theta)$$

where, y is class variable and X is a dependent feature vector (of size n) where:

$$\Theta = (\Theta\, 1, \Theta\, 2, \ldots .\, \Theta\, n)$$

Ex. Θ = (Rainy, Hot, High, False)

ci = No

So basically, $P(\Theta\, |Ci)$ here means, the probability of "Not playing golf" given that the weather conditions are "Rainy outlook", "Temperature is hot", "high humidity" and "no wind".

Naive Bayes classification

Put a naive assumption to the Bayes' theorem, which is, **independence** among the features. So now, we split **evidence** into the independent parts.

$$P(c_i|\theta_1, \theta_2, \ldots, \theta_n) = \frac{P(\theta_1, \theta_2, \ldots, \theta_n|c_i)P(c_i)}{P(\theta_1, \theta_2, \ldots, \theta_n)}$$

$$P(c_i|\theta_1, \theta_2, \ldots, \theta_n) = \frac{[P(\theta_1|c_i)P(\theta_2|c_i)P(\theta_3|c_i)\ldots P(\theta_n|c_i)]P(c_i)}{P(\theta_1, \theta_2, \ldots, \theta_n)}$$

$$P(c_i|\theta_1, \theta_2, \ldots, \theta_n) = \frac{P(c_i)[\prod_{m=1}^{n} P(\theta_m|c_i)]}{P(\theta_1, \theta_2, \ldots, \theta_n)}$$

We can just calculate $P(c_i|\theta_1, \theta_2, \ldots, \theta_n)$ for all c_i and the class prediction is the c_i with maximal value of $P(c_i|\theta_1, \theta_2, \ldots, \theta_n)$

This classifier then is called **Naive Bayes**. In formal from, we can write as follows.

$$class = \underset{c_i}{argmax} \frac{P(c_i)[\prod_{m=1}^{n} P(\theta_m|c_i)]}{P(\theta_1, \theta_2, \ldots, \theta_n)}$$

Naive Bayes classification

And since the denominator for all a c_i same, we can simplify as follows.

And **this is our final Naive Bayes classifier**.

$$\boxed{class = \underset{c_i}{argmax} P(c_i)[\prod_{m=1}^{n} P(\theta_m|c_i)]}$$

So, finally, we are left with the task of calculating P(ci) and P(Θ m | ci).

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(Θ m | ci) for each Θ m $_i$ in Θ and ci in c. All these calculations have been demonstrated in the tables below:

Naive Bayes classification

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%

Naive Bayes classification

For example, probability of playing golf given that the temperature is cool, i.e P(temp. = cool | play golf = Yes) = 3/9.

Also, we need to find class probabilities (P(c)) which has been calculated in the table 5.

For example, P(play golf = Yes) = 9/14.

so now, we are done with our pre-computations and the classifier is ready!
Let us test it on a new set of features (let us call it today):

today = (Sunny, Hot, Normal, False)

Naive Bayes classification

So, probability of playing golf is given by:

$$P(Yes|today) = \frac{P(SunnyOutlook|Yes)P(HotTemperature|Yes)P(NormalHumidity|Yes)P(NoWind|Yes)P(Yes)}{P(today)}$$

and probability to not play golf is given by:

$$P(No|today) = \frac{P(SunnyOutlook|No)P(HotTemperature|No)P(NormalHumidity|No)P(NoWind|No)P(No)}{P(today)}$$

Since, P(today) is common in both probabilities, we can ignore P(today) and find proportional probabilities as:

$$P(Yes|today) \propto \frac{2}{9}.\frac{2}{9}.\frac{6}{9}.\frac{6}{9}.\frac{9}{14} \approx 0.0141$$

and

$$P(No|today) \propto \frac{3}{5}.\frac{2}{5}.\frac{1}{5}.\frac{2}{5}.\frac{5}{14} \approx 0.0068$$

Now, since

$$P(Yes|today) + P(No|today) = 1$$

These numbers can be converted into a probability by making the sum equal to 1 (normalization):

$$P(Yes|today) = \frac{0.0141}{0.0141+0.0068} = 0.67$$

and

$$P(No|today) = \frac{0.0068}{0.0141+0.0068} = 0.33$$

Since

$$P(Yes|today) > P(No|today)$$

So, prediction that golf would be played is 'Yes'.