Naive Bayes Classifier

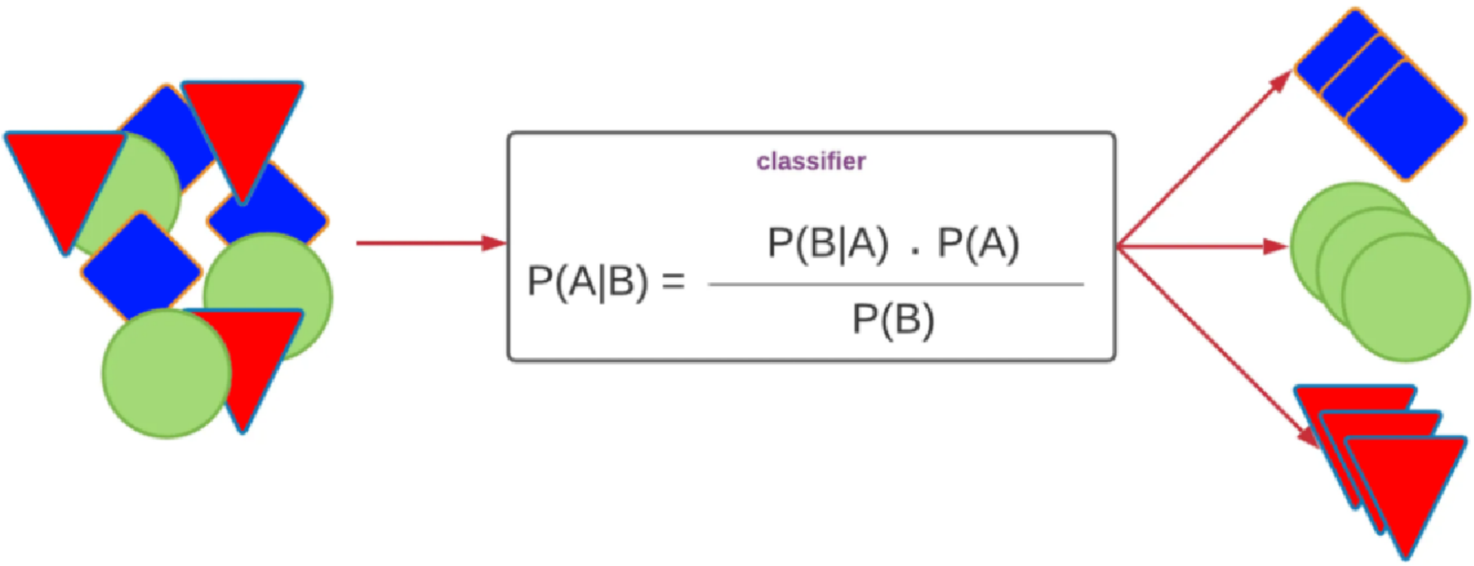

Table Of Contents

Naive Bayes classifier

→ If we have a dataset, where relationship with one single feature may effect the output/label, then we cannot generalize about all the features at a time!!

Eg:-

1) The food I ordered is good.
2) The food I ordered is bad.

Here only the last word signifies about the sentiment of statement.

(*) This is where Naive Bayes helps in.

→ Naive Bayes will not try to generalize complete features at a time and instead tries to build a independent relation b/w each & every input feature with output label, so that we get a better idea.

How it works?

→ Naive Bayes is not a single algorithm, but a family of algorithms where all of them share a common principle, i.e; they are based on "Bayes

→ Bayes' theorem finds the probability of an event occuring given the probability of another event that has already occured.

mathematically:

$$\boxed{P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)}}$$

where A & B are events & $P(B) \neq 0$

→ Basically we are trying to find the Probability of A, given the event B is true. Event B is also termed as evidence.

→ P(A) is the Priori of A, that is the probability of event before evidence is seen (Prior Probability).
Here, event B occurance

→ P(A/B) is the posteriori probability, that is the probability of event after evidence is seen.

Eg: lets take a dataset of Playing golf based on climate.

S. No	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

→ Here we can say that if it's sunny, mostly likely a person wouldn't go for game. Similar case if it's too windy / too rainy.

→ So we can clearly see that a single feature will affect the label directly. Hence, we will build model using independent relationship of each feature with output label by leveraging Naive Bayes.

→ But before we apply Naive Bayes, we should make sure it follows the assumptions.

☆ ⓘ The fundamentals Naive Bayes Assumption is that each feature makes an <u>independent</u>, and <u>equal</u> contribution to the outcome.

→ And we can clearly say no pair of features are dependent, Eg- the temperature being hot has nothing to do with humidity / being rainy. Hence, features are independent.

→ Knowing only temperature & humidity alone can't predict the outcome, hence assumed all features contribute equally to outcome.

we have, Bayes' Theorem as,

$$\boxed{P(y/x) = \frac{P(X|y) \cdot P(y)}{P(X)}}$$

y - class variable

X - dependent features.

→ And by taking naive assumption to bayes theorem, which is independence among the features, we get $\boxed{P(A,B) = P(A) \cdot P(B)}$

$$P(y \mid (x_1, x_2, \cdots x_n)) = \frac{P((x_1, x_2, \cdots x_n)/y) \cdot P(y)}{P(x_1, x_2 \cdots x_n)}$$

Based on Naive assumption of independence,

$$= \frac{[P(x_1 \mid y) \cdot P(x_2 \mid y) \cdots P(x_n \mid y)] \cdot P(y)}{P(x_1) \cdot P(x_2) \cdots P(x_n)}$$

$$= \frac{P(y) \cdot \prod_{i=1}^{n} P(x_i \mid y)}{\prod_{i=1}^{n} P(x_i)}$$

constant for any class variable in a class.

→ Now, To make a classifier model, we find the Probability of a given set of inputs for all possible values of the class variable y and pick up the output with maximum probability.

$$\therefore \boxed{y = \text{argmax}_y \, P(y) \prod_{i=1}^{n} P(x_i \mid y)}$$

→ so, to solve this problem we need to calculate $P(y)$ & $P(x_i \mid y)$ → class probability, Conditional Probability.

→ so now we calculate $P(\text{Sunny} \mid \text{yes})$, $P(\text{overcast} \mid \text{yes})$, $P(\text{Rainy} \mid \text{yes})$, $P(\text{Sunny} \mid \text{No})$, $P(\text{overcast} \mid \text{No})$, $P(\text{Rainy} \mid \text{No})$, similarly for each feature. $\boxed{P(x_i \mid y_j)}$

we get,

Out look Features

	Y	N	P(Y)	P(N)
Sunny	3	2	3/9	2/5
overcast	4	0	4/9	0/5
Rainy	2	3	2/9	3/5
Total	9	5	100%	100%

Temperature Feature

	Y	N	P(Y)	P(N)
Hot	2	2	2/9	2/5
mild	4	2	4/9	2/5
cold	3	1	3/9	1/5
Total	9	5	100%	100%

Humidity Feature

	Y	N	P(Y)	P(N)
High	3	4	3/9	4/5
Normal	6	1	6/9	1/5
Total	9	5	100%	100%

Wind Feature

	Y	N	P(Y)	P(N)
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%

→ Here we calculated $P(x_i | y_j)$ for each x_i & y_j, And also $P(y_j)$ for each class variable in y.

Prediction

→ Now, if we have a new datapoint as

(Sunny, Hot, Normal, False) = today climate.

$$P(y \mid \text{today climate}) = \frac{P(\text{Sunny}|\text{yes}) \cdot P(\text{Hot}|\text{yes}) \cdot P(\text{Normal}|\text{yes}) \cdot P(\text{False}|\text{yes}) \cdot P(\text{yes})}{P(\text{today})}$$

$$P(\text{yes} \mid \text{today}) \propto \frac{3}{9} \cdot \frac{2}{9} \cdot \frac{6}{9} \cdot \frac{6}{9} \cdot \frac{9}{14} \approx 0.0211$$

$P(N|today) \propto \frac{2}{5} \cdot \frac{3}{5} \cdot \frac{1}{5} \cdot \frac{2}{5} \cdot \frac{5}{11} \approx 0.1142$

However, we know,

$$P(y|today) + P(N|today) = 1$$

Apply reciprocal,

$$1 = \frac{1}{P(y|today) + P(N|today)}$$

∴ multiply $P(y|today)$

$$P(y|today) = \frac{P(y|today)}{P(y|today) + P(N|today)}$$

$$P(y|today) = \frac{0.024}{0.0211 + 0.1142} = \frac{0.024}{0.1353} = 0.155$$

$$P(N|today) = \frac{0.1142}{0.0211 + 0.1142} = \frac{0.1142}{0.1353} = 0.844.$$

So, clearly $\boxed{P(N|today) > P(y|today)}$

Hence, Play golf → No for today's climate.

→ However, this above method is for the categorical data.

→ And in case of continuous data we need to make assumptions regarding the distribution of values of each feature.

→ ☆ The different Naive Bayes classifier differ mainly by the assumptions we make regarding the distribution of $P(x_i|y)$

⊙ Gaussian Naive Bayes classifier :-

→ If the feature variables are a continuous variables, then we cannot use the Naive Bayes, and instead we should be using Gaussian Naive Bayes Classifier.

→ In Gaussian Naive Bayes, continuous values associated with each feature are assumed to be distributed according to a Gaussian distribution, also called "Normal" distribution.

→ The likelihood of the features is assumed to be Gaussian, hence conditional probability is,

$$\boxed{P(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \cdot \exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right)}$$

Other popular Naive Bayes classifier are :-

⊙ Multinomial Naive Bayes :- Feature vectors represent the frequencies with which certain events have been generated by multinomial distribution. This is typically used for document classification.

⊙ Bernoulli Naive Bayes :- Assumes that the features are binary or categorical. It is particularly useful for document classification like MNB, and where binary term occurrence features are used rather than term frequencies.