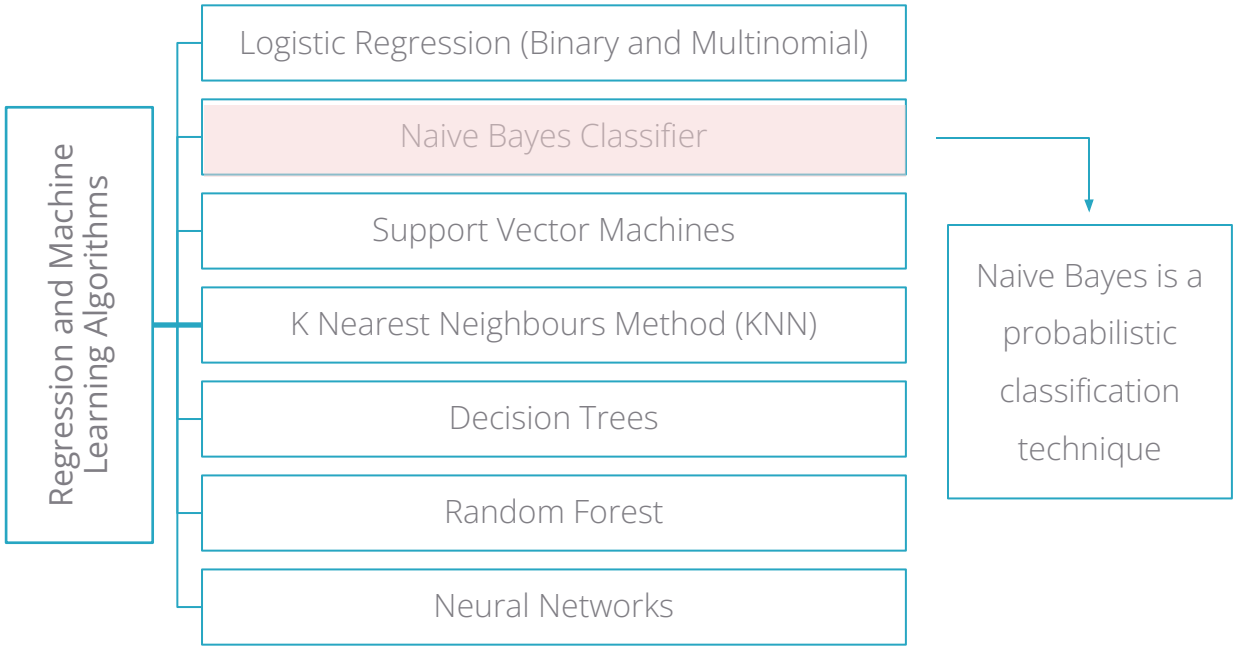
Naive Bayes Classifier - I

Classifier Based on Bayes' Theorem

Contents

Classification Methods

Apart from logistic regression, several types of machine learning algorithms are effective in classification and prediction.



About Naive Bayes Classifier

- Simple **probabilistic classifier based on Bayes Theorem.**
- It can be used as an **alternative method to logistic regression (Binary or Multinomial).**
- It **assumes conditional independence among the predictors.**
- It is particularly **suited when the dimensionality of the inputs is high.**

Despite its simplicity, Naive Bayes can often outperform more sophisticated classification methods.

Conditional Probability

The conditional probability of an event B is the probability that event B will occur given the knowledge that an event A has already occurred. This probability is written as $P(B|A)$.

- If A and B are independent events then

$$P(B|A) = P(B)$$

- An unbiased die, with numbers 1-6 is tossed

  A: Getting a number greater than 1

  B: Getting an even number

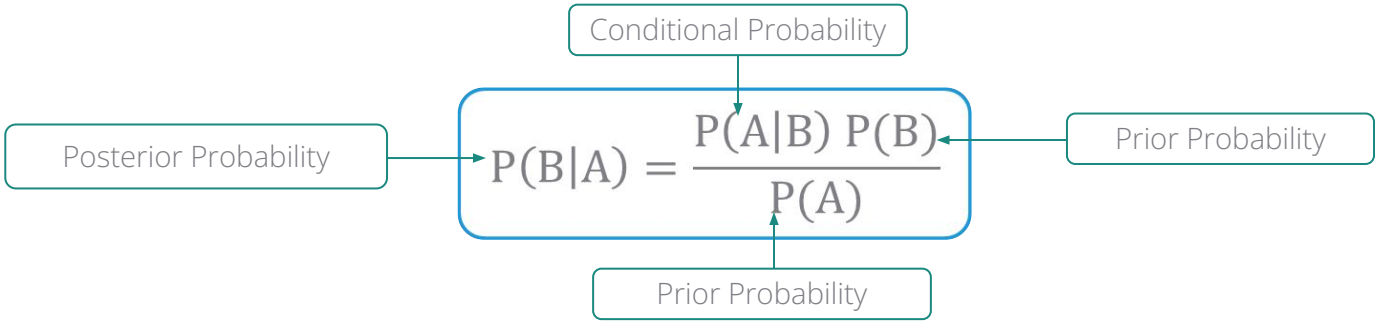
$$P(A) = 5/6$$

$$P(B) = 3/6$$

$$P(B|A) = 3/5$$

Here the sample space has 5 points given A has occurred.

Bayes Theorem



where

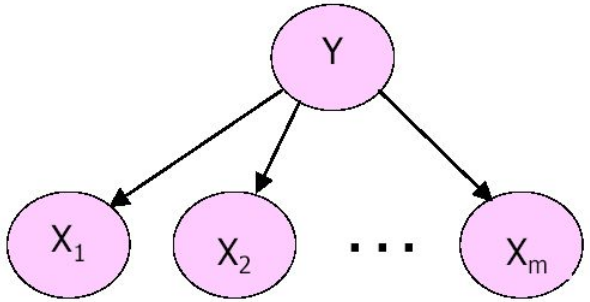
P(A) : Prior probability or marginal probability of A

P(A | B) : Conditional probability of A given B

P(B | A) : Conditional probability of B given A

P(B) : Prior or marginal probability of B

Naive Bayes Framework



Y : Categorical Dependent Variable

X_i : Categorical/Continuous Independent Variable

Objective: To estimate Y given the values of X_i's or

To estimate $P(Y|X_1, X_2, \ldots, X_m)$ using the Naïve Bayes Classifier

Assumption: All X_i's are conditionally independent of each other

Naive Bayes Framework - Example

Consider a simple example where Y is binary (response to a certain question) with 2 independent categorical variables X_1 and X_2

We classify	Y = 1 "Buyer" Y = 0 "Non-Buyer"
Let $\mathbf{X_1}$ denote **age of the individual**	X_1 = 0 for age group 25-30 years X_1 = 1 for age group 31-40 years
Let $\mathbf{X_2}$ denote **gender**	X_2 = 0 if Gender=female X_2 = 1 if Gender=male

Classification Rule

For the given values of X_1 and X_2 we want to know if the individual will be a potential buyer or not. Using Naive Bayes classifier we estimate:

$$P(Y = 0|X_1 = a_1, X_2 = a_2)$$

&

$$P(Y = 1|X_1 = a_1, X_2 = a_2)$$

where a_1 and a_2 are values of X_1 and X_2 for a particular respondent

We classify Y = 0 if $P(Y = 0|X_1 = a_1, X_2 = a_2) > 0.5$ OR

Y = 1 if $P(Y = 1|X_1 = a_1, X_2 = a_2) > 0.5$

In the general case i.e. when Y has more than 2 categories we compare $P(Y= y_k \mid X)$ for all values of y_k and classify $Y = y_k$ for which $P(Y= y_k \mid X)$ is the maximum

Expected Output

Once the classification rule is applied the output can be shown as follows:

Case#	X1	X2	$P(Y=1/X_1,X_2)$	$P(Y=0/X_1,X_2)$	Y classified as
1	1	0	0.44	0.56	0
2	1	1	0.7	0.3	1
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
240	0	0	0.2	0.8	0

Advantages of Naive Bayes Method

- Classification rule is simple to understand.
- The method requires a small amount of training data to estimate the parameters necessary for classification.
- The evaluation of the classifier is quick and easy.
- The method can be a good alternative to logistic regression.

Limitations of Naive Bayes Method

- Assumption of conditional independence of the independent variables is highly impractical.
- In case of continuous independent variables the density function must be known or assumed to be normal.
- In case of categorical independent variables the probabilities cannot be calculated if the count in any conditional category is zero. For instance: If there are no respondents in the age group 25-30 yrs. then $P(X_1=0 \mid Y=1) = 0$

?

How to deal with such cases?

If a category has zero entries we replace 0 by 0.5/n (n = sample size) so that the probability expression does not reduce to zero.

Quick Recap

In this session, we learnt **Naive Bayes Classification** technique:

Conditional Probability and Bayes' Theorem

- The conditional probability of an event B is the probability that event B will occur given the knowledge that an event A has already occurred.
- $P(B|A) = P(A|B)\,P(B) \,/\, P(A)$

Naive Bayes Classifier

- To estimate Y given the values of X_i's or $P(Y|X_1, X_2, \ldots, X_m)$ using the Naïve Bayes Classifier.
- **Assumption:** All X_i's are conditionally independent of each other.
- **Advantages:** Simple classification rule, requires a small amount of training data to estimate the parameters necessary for classification, Evaluation of the classifier is quick and easy, Good alternative to logistic regression.
- **Major drawback:** Assumption of conditional independence of the independent variables is highly impractical.