Naive Bayes

Mannem Charan AI21BTECH11019

Abstract

This report consists of my basic understanding of one of the popular Ml methods "Naive Bayes".

1 Naïve Bayes

Naïve Bayes is a supervised learning algorithm which is predominantly used for classification problems. It is simple and most effective method out their to solve classification problems (even effective for multi-class classification).It uses probabilistic technique for constructing classifiers.It assumes that the occurence or non- occurence of a particular feature independent of other features.For example,let say we want to claasify a fruit, the normal features that we take are color, size and shape.In Naïve Bayes the color of the fruit has nothing to do with size of fruit , similarly shape.It may be unreaslitic for real data but there are array of complex problems which can effectively solved by Naïve Bayes.

2 What is the name Naïve Bayes means?

- **Naïve:**The word Naïve comes from it's assumption that the parameters involved in input variable are independent to each other.Even though it is bit naïve to assume that,it doesnot do any harm in predicting the classes.Think of it in this way, if features are dependent like outlook of the day and humidity,even though both are correlated,taking both as independent gives us a double-evidence intuitively even though it is bit wrong to do.
- **Bayes:** The word Bayes is given since the model is mainly based on Bayes theorem.To understand Bayes theorem, let us take A as hypothesis and B as evidence. Then from Bayes theorem we can say that,

$$\Pr(A|B) = \frac{\Pr(A)\Pr(B|A)}{\Pr(B)} \tag{2.1}$$

3 Understanding Naïve Bayes

With the assumption in the hand, Naïve Bayes try to predict the conditional probabilities of each class given the data point.To understand that let us say X is the input variable with features x_1,x_2,x_3,..,x_k and let say y is the target attribute. Now we want to

find the probability that y is the label of the data point given the data point X,i.e., we want to the conditional probability,

$$\Pr(y|\mathbf{X}) \tag{3.1}$$

Now from Bayes theorem we can write that,

$$\Pr(y|\mathbf{X}) = \frac{\Pr(y)\Pr(\mathbf{X}|y)}{\Pr(\mathbf{X})} \tag{3.2}$$

Now from our assumption we simplify it as,

$$= \frac{\Pr(y)\Pr(x_1|y)\Pr(x_2|y)\Pr(x_3|y)\ldots\Pr(x_k|y)}{\Pr(x_1)\Pr(x_2)\Pr(x_3)\ \ldots\Pr(x_k)} \tag{3.3}$$

$$= \frac{1}{Z}\Pr(y)\prod_{i=1}^{k}\Pr(x_i|y) \tag{3.4}$$

where $Z = \prod_{i=1}^{k}\Pr(x_i)$, the value of it is constant since it depends on occurence of features in the data.And with equation (3.4), we try to predict whether the given input vector is assigned with the label y or not. And if the label like let say color of fruit has more outcomes (red, orange, blue etc) possible,in that case we have to assign a class label $\hat{y}$ to X which has maximum conditional probability, the same thing mathematically can be wriiten as ,

$$\hat{y} = \underset{y}{\operatorname{argmax}}\,\Pr(y)\prod_{i=1}^{k}\Pr(x_i|y) \tag{3.5}$$

Using the above function we will assign class labels to each data point.

4 Different flavours of Naïve Bayes

There are different types Naïve Bayes classifiers based on the distributions taken by the predictors/features.

1) **Multinomial Naïve Bayes:** In this classifier, each $\Pr(x_i|y)$ takes multinomial distribution.It is generally used in document classification, like reviewing a text as a email/news article.It takes features/predictors as the words/tokens in the text with frequency in the document as it's values.
2) **Binomial Naïve Bayes:** Here the predictors/features will only take two values(basically yes or no).Since it is similar to binomial distribution it is named as "Binomial Naïve Bayes" model.
3) **Gaussian Naïve Bayes:** Often it is possible that the predictors/features take continuous values.So we assume that these values asssociated with each class follow

gaussian distribution,i.e.,

$$\Pr(x_i|y) = \frac{1}{2\pi\sigma_y^2} \exp - \frac{\left(x_i - \nu_y\right)^2}{2\sigma_y^2} \tag{4.1}$$

5 Advantages of Naïve Bayes

1) It is one of the fast and easy ML algorithms to predict the class of a dataset
2) It is effective for binary and even multiclass classification.
3) It is predominantly used in Natural Language processing(*NLP*).

6 Disadvantages of Naïve Bayes

1) It is stupid to believe that the features are independent.
2) Not effective for regression problems.

7 Questions

1) What is naïve about Naïve Bayes classifier?
2) What is the main principle behind Naïve Bayes classifier?
3) What are the types of Naïve Bayes Classifiers?
4) What we will do in training phase of Naïve Bayes model?
5) Where we commonly use Naïve Bayes?

8 Answers

1) It is the assumption that the features are independent.As in real life mostly the features are dependent on each other.
2) The main principle used in Naïve Bayes is the Bayes theorem.
3) Refer section 4.
4) In training phase, you will collect whatever information you need to calculate $\Pr(x_i|y)$ and $\Pr(y)$
5) It is commonly used in NLP to give tags to the text.