Naive Bayes

A Naive Bayes classifier is a probabilistic machine learning model that’s used for classification task. The classifier is based on the Bayes Theorem.

The Bayes’ Theorem is based on conditional probability or in simple terms, the likelihood that an event (A) will happen given that another event (B) has already happened. The Bayes’ Theorem assumes that features are statistically independent.

$$\underbrace{P(c \mid x)}_{\text{Posterior Probability}} = \frac{\overbrace{P(x \mid c)}^{\text{Likelihood}}\,\overbrace{P(c)}^{\text{Class Prior Probability}}}{\underbrace{P(x)}_{\text{Predictor Prior Probability}}}$$

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Above equation can be written as

Posterior = (Likelihood * Prior) / Evidence

c - The class variable

x - Features i.e. x_1, x_2….x_n.

P(c|x) = Posterior

P(x|c) = Likelihood

P(c) = Prior

P(x) = Evidence

Types

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.

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)}$$

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.

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)}$$

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. Since the way the values are present in the dataset changes, the formula for conditional probability changes to

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

Example

Let’s say we have data on 1000 pieces of fruit. The fruit being a Banana, Orange or some other fruit and imagine we know 3 features of each fruit, whether it’s long or not, sweet or not and yellow or not.

Fruit	Long	Sweet	Yellow	Total
Banana	400	350	450	500
Orange	0	150	300	300
Other	100	150	50	200
Total	**500**	**650**	**800**	**1000**

- Compute the ‘Prior’ probabilities for each of the class of fruits.

That is the proportion of each fruit class out of all the fruits from the population. You can provide the ‘Priors’ from prior information about the population. Otherwise, it can be computed from the training data.

Let’s compute from the training data. Out of 1000 records in training data, you have 500 Bananas, 300 Oranges and 200 Others.

P(Y=Banana) = 500 / 1000 = 0.50

P(Y=Orange) = 300 / 1000 = 0.30

P(Y=Other) = 200 / 1000 = 0.20

So the respective priors are 0.5, 0.3 and 0.2.

- Compute the probability of evidence that goes in the denominator.

This is nothing but the product of P of Xs for all X. This is an optional step because the denominator is the same for all the classes and so will not affect the probabilities.

P(x1=Long) = 500 / 1000 = 0.50

P(x2=Sweet) = 650 / 1000 = 0.65

P(x3=Yellow) = 800 / 1000 = 0.80

- Compute the probability of likelihood of evidences that goes in the numerator.

Here I have done it for Banana alone. Probability of Likelihood for Banana

P(x1=Long | Y=Banana) = 400 / 500 = 0.80

P(x2=Sweet | Y=Banana) = 350 / 500 = 0.70

P(x3=Yellow | Y=Banana) = 450 / 500 = 0.90

So, the overall probability of Likelihood of evidence for Banana = 0.8 * 0.7 * 0.9 = 0.504

Similarly, you can compute the probabilities for ‘Orange’ and ‘Other fruit’. The denominator is the same for all 3 cases, so it’s optional to compute

- Substitute all the 3 equations into the Naive Bayes formula, to get the probability that it is a banana.