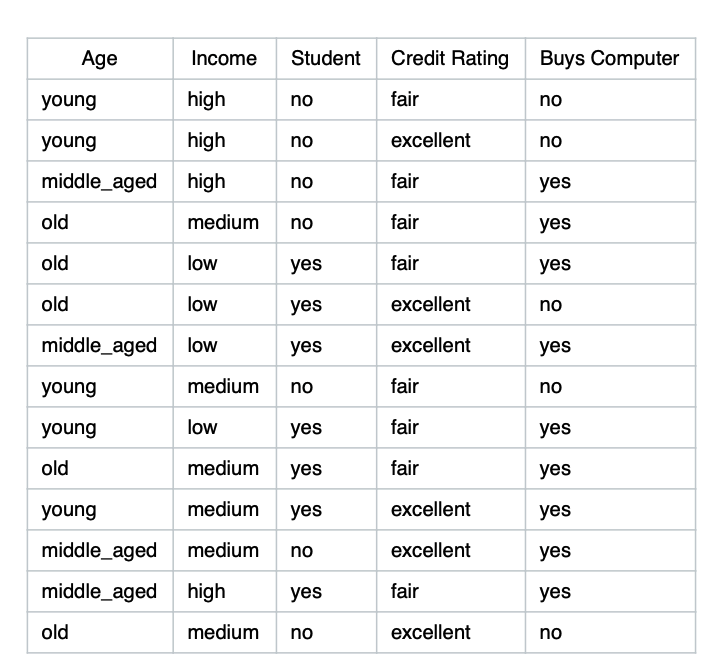
Gaussian Naive Bayes

Before we start let’s remember what is naive bayes is

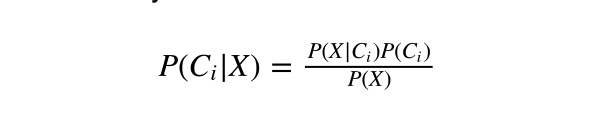
Naive Bayes

The naive bayes classifier performs probabilistic prediction based on the probability values of class membership.

The\_tech shops data shows that 5 out of 15 customers purchased computers and 10 out of 15 did not buy a computer. The probability of the next customer buying a computer would be 5/15. This is a simple probability. However naive bayes consider prior probability and record attributes. Following data shows more features about customers.



Instead of making simple predictions by using simple probability Naive Bayes consider customers prior conditions.



X represents attributes and X are the labels.

A customer came in with futures

X(age=young, income = medium, student = yes, Credit Rating = fair). In this case the above formula will give the probability of a class. The prediction will be based on higher probability.

9y , 5 n

P(C1) = 9/14 the probability of yes

P(C2) = 5/14 the probability of no

P(x/C1) = P(young/yes) \* P(medium/yes) \* P(student/yes) \* P(fair/yes)

= 2/9 \* 4/9 \* 6/9 \* 6/9 = 0.044

p(x/c2) = P(young/no) \* P(medium/no) \* P(student/no) \* P(fair/no)

= 3/5 \* 2/5\* 1/5\* ⅖ = 0.0019

Since they share the same denominator we can just compare them

P(C1) > P(C2) therefore we conclude that customer x will belong to class yes.

Advantages

Work quickly and can save a lot of time

Naive bayes are suitable for solving multi class prediction problems.

If its assumptions of the independence of features holds true, it can perform better than other methods and requires less training data

Naive bayes are better for categorical input variables than numerical variables.

Not biased to outliers

Disadvantages

Naive Bayes assumes all features are independent, rarely happening in real life.

Naive Bayes deals with categorical data whereas Guassian Naive Bayes deals with continuous variables.

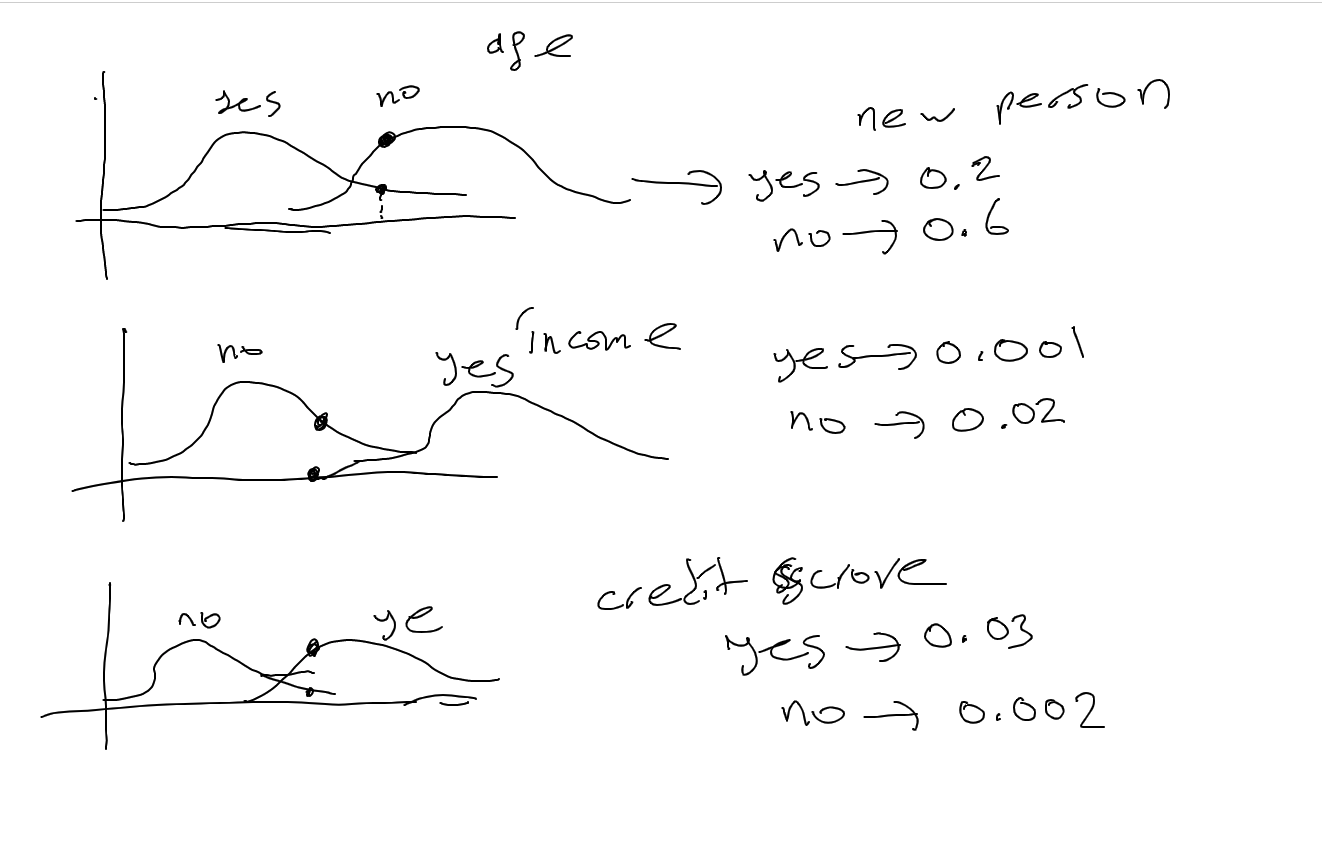
|  |  |  |  |
| --- | --- | --- | --- |
| age | income | Credit score | Buy computer |
| 24 | 34000 | 670 | y |
| 12 | 49999 | 700 | y |
| 23 | 565656 | 720 | y |

Let’s say there 20 record and 10 y and 10 n class labels

First we need to find prior probabilities

Y = 10/20 , n= 10/20

Then we need to computer mean and standard deviation of each attributes for yes and no graph normal distributions



Then a new person comes in with the dot futures.

log(0.5\*0.2 … ) = log(0.5) + log(0.2)+..

p(y) = log(0.5) + log(0.2) + log(0.001) + log(0.03) = -12.7

p(n) = log(0.5) + log(0.6) +llog(0.02) +log(0.002) = -11.3

Since p(n) greater new person will belong to class no