

Gaussian Naive Bayes is a simple probabilistic classifier that uses Bayes theorem with independent assumptions between the features to classify data. Let, the classifier has m elements which is denoted with $X = X_1, X_2, \dots, X_m$ and n classes which is denoted with $C = c_1, c_2, \dots, c_n$. The Bayes theorem is stated in following equation,

$$P(C_i | X_j) = \frac{P(X_j | C_i) * P(C_i)}{P(X_j)}$$

Here,

C_i = Denotes the class

X_j = Denotes a single featured element

$P(A | B)$ = Denotes the probability of observing A after B is observed.

$P(A)$ = Denotes the probability of observing A

For multiple feature the equation is changed to,

$$P(C_i | x_1, x_2, x_3, \dots, x_n) = \frac{P(x_1 | C_i) * P(x_2 | C_i) * \dots * P(x_n | C_i)}{P(X_j)}$$

Here,

$x_1, x_2, x_3, \dots, x_n$ are features of X_j .

Gaussian Naive Bayes is really fragile to over fitting without any regularization assumption. Also, it is based on naive assumptions that are not generally concordant with the data.

1991-1992