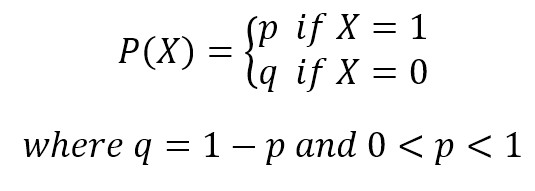
**Bernoulli naive Bayes**

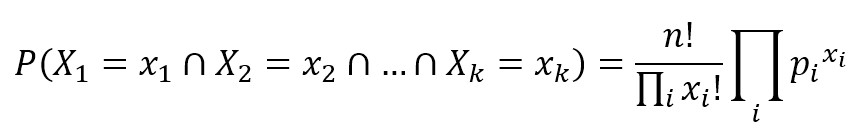
If X is random variable Bernoulli-distributed, it can assume only two values (for simplicity, 0 and 1 is used ) and their probability is:



The binomial model is useful if the feature vectors are binary (i.e., 0s and 1s). One application would be text classification with a bag of words model where the 0s 1s are "word occurs in the document" and "word does not occur in the document"

**Multinomial naive Bayes**

A multinomial distribution is useful to model feature vectors where each value represents, for example, the number of occurrences of a term or its relative frequency. If the feature vectors have *n*elements and each of them can assume *k*different values with probability *pk*, then:

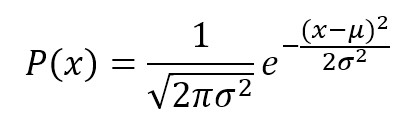


The conditional probabilities P(xi|y) are computed with a frequency count (which corresponds to applying a maximum likelihood approach). It’s possible to assign all non-negative values, however, larger values will assign higher probabilities to the missing features and this choice could alter the stability of the model.

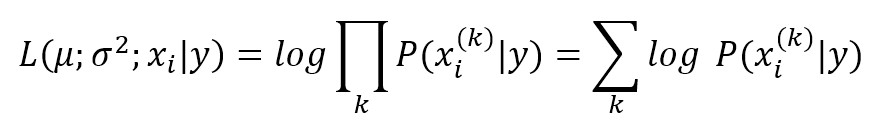
The multinomial naive Bayes model is typically used for discrete counts. E.g., if we have a text classification problem, we can take the idea of bernoulli trials one step further and instead of "word occurs in the document" we have "count how often word occurs in the document", you can think of it as "number of times outcome number x\_i is observed over the n trials"

**Gaussian Naive Bayes**

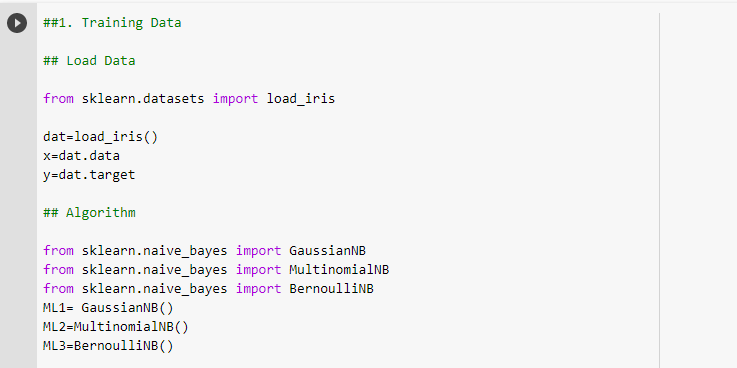
Gaussian Naive Bayes is useful when working with continuous values which probabilities can be modeled using a Gaussian distribution:



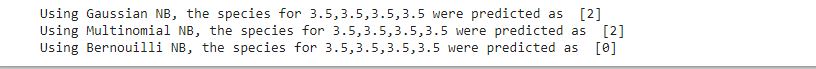
The conditional probabilities P(xi|y) are also Gaussian distributed and, therefore, it’s necessary to estimate mean and variance of each of them using the maximum likelihood approach.



The parameters σ and μ are estimated using maximum likelihood.







Here in the Iris dataset, the features are of real valued data and continuous. Hence GaussianNB is more suitable for classification and BernouilliNB performed worse as seen in cross validation and gives incorrect result, ie, [0].