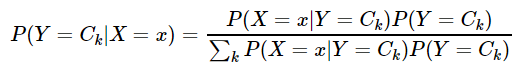
Considering that reviews are unstructured data, it’s hard to directly apply them to our model. Hence, we try to use sentiment analysis to convert them into structured data. First, we need to select features that most related to sentiment of the reviews. We use a simple feature extractor that indicates which words in the training set are contained in a document provided by a package. Then, we need to use a classification algorithm that performs well. After a set of testing and comparison, we finally choose Naive Bayesian classifier which will perform well if features are independent. Naive Bayesian classifier is based on Bayes' theorem. It assumes that every feature exploited by the model is independent. Hence, the possibility of y being equal to Ck when we are given a vector X0 is



比较画图