**Bernoulli Naive Bayes:**Itassumes that all our features are binary such that they take only two values. Means **0s** can represent “word does not occur in the document” and **1s** as "word occurs in the document”.

**Multinomial Naive Bayes**: It is used when we have **discrete data** (e.g., movie ratings ranging 1 and 5 as each rating will have certain **frequency** to represent). In text learning we have the count of each word to predict the class or label.

**Gaussian Naive Bayes:**Because of the assumption of the **normal distribution**, Gaussian Naive Bayes is used in cases when all our features are **continuous**. For example, in **Iris dataset** features are sepal width, petal width, sepal length, petal length. So, its features can have different values in data set as width and length can vary. We can’t represent features in terms of their occurrences. This means data is continuous. Hence, we use Gaussian Naive Bayes here.