Bernoulli Naïve Bayes: It assumes that all our features are binary such as 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 Naïve Bayes: It is used when we have **discrete data** (e.g. movie ratings ranging 1 to 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 Naïve Bayes: Because of the assumption of the **normal distribution,** Gaussian Naïve 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 Naïve Bayes here.