Naive Bayes Algorithm (allows the machine to take decisions on its own, e.g. Types of books (different people recommended it)  
It considers features as independent, they have individual weights.  
  
Generative 🡪 Naïve Bayes

Read Laplace Transformation for the basic classes.

Naïve Bayes Algorithm 🡪 Follows Bayes Theorem (Creates a different model each class)  
P(Class|Feature)=[P(Feature|Class)\*P(Class)]/(P(Features))  
P(A|B)=((B|A)\*P(A))/P(B)  
A=”Spam Email”   
B=”Free”



Classification of Models  
Generative Models🡪Naïve Bayes, as it can generate new data  
Discriminative Models  
  
To correct, the distinguishing characters between two models (crosses and circle) and then when we check for square, we give it a probability of 0.  
So, we use Laplace Smoothing. It adds a probability of 1 for those which has something wrong (has probability 0).



We are adding 1, when we encounter 0, so that we can regularize our curve



Generative Advisory Neural Network (Read from Google)

Types of Naïve Bayes  
Gaussian Naïve Bayes 🡪 One type of data is coming in continuously. Eg age, temperature, etc  
Multinomial Naïve Bayes 🡪eg to see the word count  
Bernoulli Naïve Bayes 🡪 presence, absence or true, false or 0 or 1

It puts (or assigns) the free to a particular value and then checks whether that particular value is coming again or not

Accuracy🡪Overall Correctness of the mode🡪

Recalls(False Negative)🡪Correct positive over actual positive🡪  
Precision(False Positive)🡪Correct Positive over total positive🡪  
F1 Score 🡪

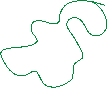
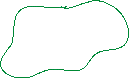
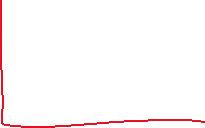
Naïve Bayes  
Text Classification, large Dataset, real time prediction

The machine is learning when and where the spams are coming and then again classifying on the basis of that

When will it be the best to avoid Naïve Bayes  
We avoid using in places where there exists high corelation

Unsupervised Data  
It is doing clustering on its own

No of clusters would be reduced .Now



Clustering woud become more prominent



DB Scan(would be taught in Day5)

Read agglomerative clustering on own







We are making a new centroid (with the coordinates of the mean of the points), and then we keep doing it iteratively till we get our perfect result. It will only work properly for 2D data.

Randomly we are assigning data points, then we are taking their man, we then make a centroid then we take the mean and take the Euclidian distance, we then repeat the process iteratively. After that, we will keep on repeating it till the no of classes in the present and the net are not the same.

Elbow Method

Disadvantages of K-Mean  
Outliers