**HMM:**

Hidden Markov Model’s: Is an un-supervised technique.

The model assumes that the system/data on which it’s being modeled follows a Markov process (randomness). The model tries to find the hidden states based on state transition probabilities (We have to specify the number of states). At first we/model start with some states probabilities (markov assumption). Each state has a probability distribution over the possible output tokens. Therefore, the sequence of tokens generated by an HMM gives some information about the sequence of states. In simple, the model starts with a known state (assumption), the next state is chosen randomly. The current state determines the probability distribution for the next state P(A|B) “So called simplest Bayesian network ” this is carried out in every step till the end. This is a iterative process, every iteration will have a state transition probability and a state. The simplest Bayesian network try to find the joint probability of all the iterations. The model performance gets better at each iteration.

**Naïve Bayes:**

A supervised learning model widely used in classification. Using Bayes theorem we can calculate posterior probability

P(c|x) = [ p(c|x) \* p(c)] /p(x)

* P(c|x) is the posterior probability of class c, given x.
* P(c) is the prior probability of class.
* P(x|c) is the likelihood which is the probability of predictor given class.
* P(x) is the prior probability of predictor.

Their 3 widely use Bayes models: Gaussian, Multinomial and Bernoulli

Gaussian assumes all its features are normally distributed, Multinomial considers the count of number times an item was repeated, Bernoulli assumes features are binominal 0 and 1.

**PCA:** Principle component analysis

PCA is a widely used dimensionality reduction technique. This techniques applies orthogonal transformation ie: linear transformation on the correlated data by finding the inner product of vectors (calculates the length of the vector and angle between the vectors) to make them linearly uncorrelated called principle components (less than actual). The first component has the highest variance, each succeeding components have the highest variance. This create a Eigen vector of co-variance matrix and eigen decomposition factorizes the matrix into a canonical form (unique vectors)