* Cluster analysis groups data based on the information found in the data that describes the objects and their relationship. The goal is that objects within a group similar to one another and different form the objects from the other groups.
* The greater the similarity within the group and larger the difference between the groups, the better the clustering is.
* If objects are labelled in known class labels it is **supervised classification** and if classification is done without any prerequisite of the data, it is called **unsupervised classification. Basic**
* K-mean clustering, aims to partition n observation into k clusters in which each of the observation belongs to the cluster with the nearest means (cluster center or cluster centroid), serving as a prototype of the cluster.
* Clusters are potential classes that captures the natural structure of the data.
* Cluster analysis provides an abstraction from individual data objects to the clusters in which those objects reside.
* Prototype of the cluster is a data object that is representative of the other objects in the cluster.
* In compression, each prototype is assigned an index, which associates with the cluster. This type of compression is known as **vector quantization.** It is often used in Image, audio, video, etc. data where many of the data are highly similar to one another, some loss of information is acceptable and a substantial reduction in the data size is desired.
* Basic algorithm of K-Mean Clustering:
* **Select the K points as the initial centroids**.
* **Repeat**
  + **Form K cluster by assigning each point to its closest centroid.**
  + **Recompute the centroid of each cluster.**
* **Until Centroid don’t change.**
* Bayes-theorem:  
  Here,  
  P(E|H) = Posterior (Probability of seeing the evidence if the hypothesis is true), P(H) = Prior (Probability a hypothesis is true before any evidence), P(E) =Probability of seeing the evidence, P(H|E) = probability a hypothesis is true given some evidence.

**P(H|E) =**