**Q1: When convention method fail?**

Aita keo parle boilo inbox a edit access dia dibo…

**Q2: When we use ROC curve? (need of it)**

The receiver operating characteristic (ROC) curve is frequently used for evaluating the performance of binary classification algorithms. It provides a graphical representation of a classifier's performance, rather than a single value like most other metrics.

Advantages by ROC curve are:

1. The curves of different models can be compared directly in general or for different thresholds.
2. The area under the curve (AUC) can be used as a summary of the model skill.
3. The shape of the curve contains a lot of information, including what we might care about most for a problem, the expected false positive rate, and the false negative rate.

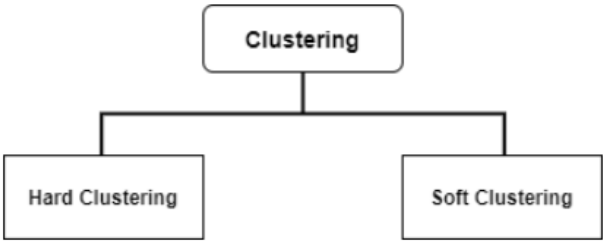
For example, if we were evaluating an email spam classifier, we would want the false positive rate to be really, really low. We wouldn't want someone to lose an important email to the spam filter just because our algorithm was too aggressive. We would probably even allow a fair amount of actual spam emails (true positives) through the filter just to make sure that no important emails were lost.

On the other hand, if our classifier is predicting whether someone has a terminal illness, we might be ok with a higher number of false positives (incorrectly diagnosing the illness), just to make sure that we don't miss any true positives (people who actually have the illness).

Additionally, ROC curves and AUC scores also allow us to compare the performance of different classifiers for the same problem.

**Q3: The difference between fuzzy C means & Kmeans clustering.**

Clustering is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups.




Hard clustering is about grouping the data items such that each item is only assigned to one cluster. **K-Means** is a famous hard clustering algorithm whereby the data items are clustered into K clusters such that each item only blogs to one cluster.

In soft clustering, instead of putting each data points into separate clusters, a probability of that point to be in that cluster assigned.
**Fuzzy C-Means** clustering is a soft clustering approach, where each data point is assigned a likelihood or probability score to belong to that cluster.

K-Means clustering and Fuzzy C-Means Clustering are very similar in approaches. Both of them are unsupervised learning algorithm. The main difference is that, in Fuzzy-C Means clustering, each point has a weighting associated with a particular cluster, so a point doesn't sit "in a cluster" as much as has a weak or strong association to the cluster, which is determined by the inverse distance to the center of the cluster. Fuzzy-C means will tend to run slower than K means, since it's actually doing more work. Each point is evaluated with each cluster, and more operations are involved in each evaluation. K-Means just needs to do a distance calculation, whereas fuzzy c means needs to do a full inverse-distance weighting.
Some real life applications of K-Means clustering: Academic performance , Diagnostic systems, Search engines, Wireless sensor networks etc.

Some real life applications of K-Means clustering: Academic performance , Diagnostic systems, Search engines, Wireless sensor networks etc. And FCM is widely applied in agricultural engineering, astronomy, chemistry, geology, image analysis etc.