Expectation Maximization

In many Machine learning problems the parameters are not known.

For example in k means clustering, clusters are not known.

We use expectation maximization to guess the parameters.

Example in k means we didn't know cluster centroids.

We also dont know the cluster values, ie which point belongs to which cluster

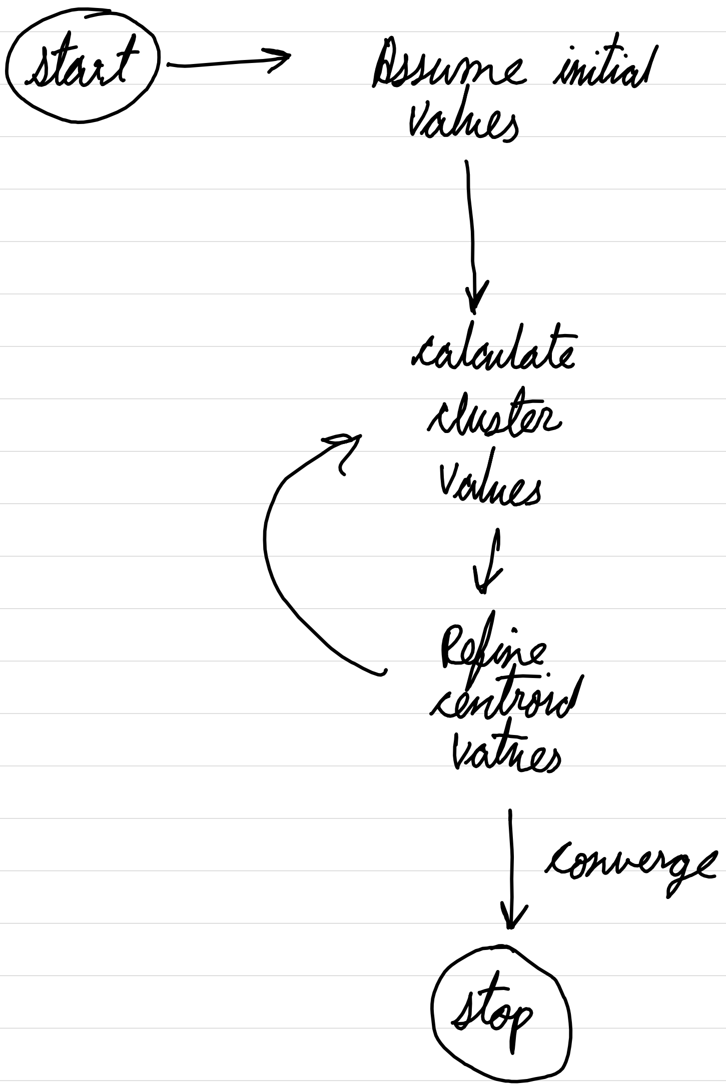
If we knew the point cluster values, we can calculate the centroids

If we had centroids then, we can calculate cluster values based on nearest neighbors

We have a chicken and egg problem

Expectation maximization is used to solve chicken and egg problem.

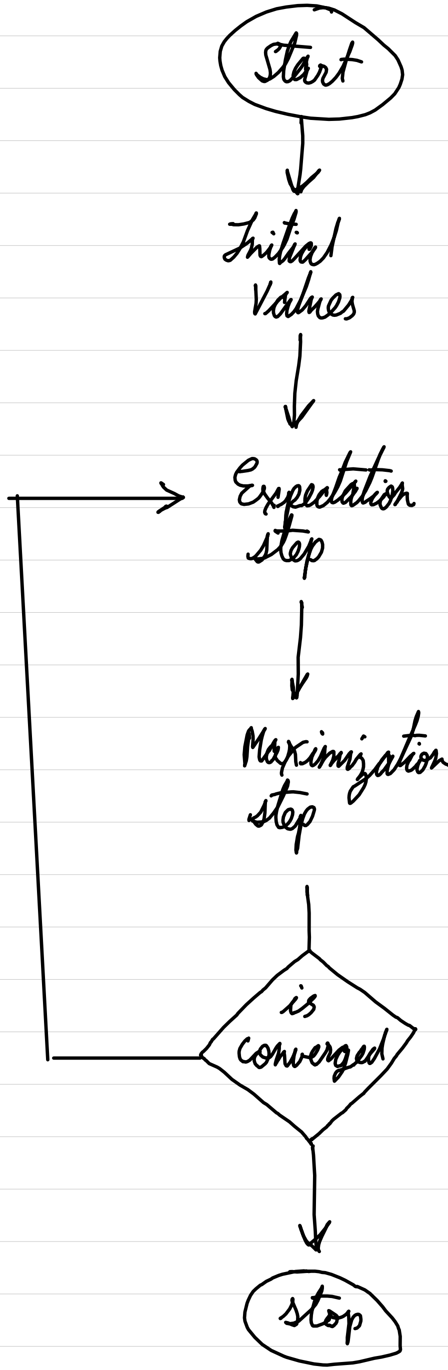
what we did in k means was



This principle is the base of Expectation maximization theorem.

Expectation Maximization works in the same way as k-means except that the data is assigned to each cluster with weights of soft probabilities

The advantage is that the model becomes generative as we define the probability distribution for each model.



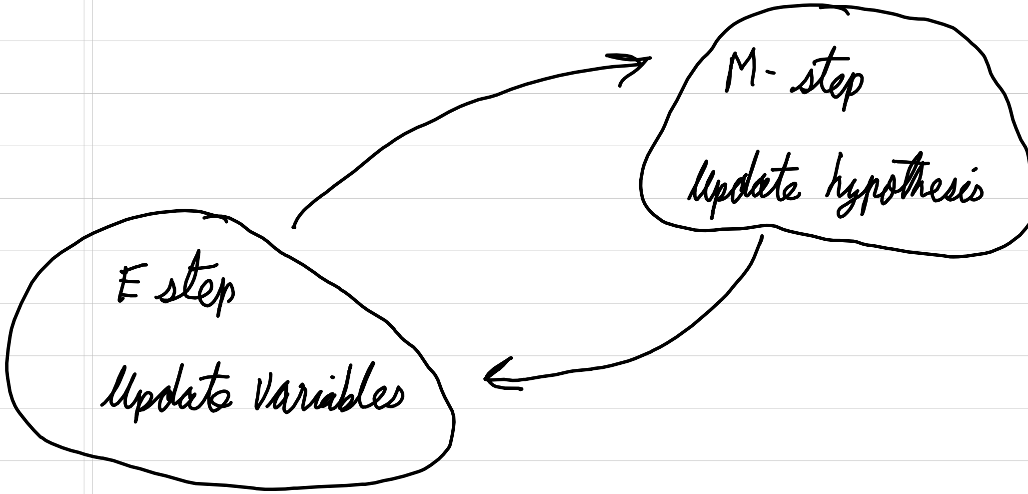
Start
Initial Values
Expectation step
Maximization step
is converged
stop

EM has two steps

E - Expectation step
M - Maximization step

1. Initialization step → Initialize with set of initial values at random
2. Expectation step → Using Observed data estimate or guess the missing data & update variables.
3. Maximization step → Use data generated in expectation step to update parameters
4. Check for convergence

At the end of expectation step, we calculate the probability of the point being in the cluster.



Iteratively refine the hypothesis & variables

It is based on the probability

Thus solve the chicken and egg problem.

This can be applied on various models most importantly gaussian Mixture Models

Used for HMM learning as well

Advantages → ① solves chicken & egg

② Easy to impliment

Dissadvantages → ① slow convergence

② local optimum

③ Time consuming if the model is large

Applications of E-M →

Used to estimate the initial values of many machine learning algorithms like HMM

GMM