DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
(ARTIFICIAL INTELLIGENCE & MACHINE LEARNING)

Expectation Maximization Algorithm

The K-means approach is an example of a hard assignment clustering, where each point can belong to only one cluster. Expectation-Maximization algorithm is a way to generalize the approach to consider the soft assignment of points to clusters so that each point has a probability of belonging to each cluster. The EM algorithm is an iterative approach that cycles between two modes. The first mode attempts to estimate the missing or latent variables, called the estimation-step or E-step. The second mode attempts to optimize the parameters of the model to best explain the data, called the maximization-step or M-step.

E-Step. Estimate the missing variables in the dataset.
M-Step. Maximize the parameters of the model in the presence of the data.

The EM algorithm can be applied quite widely, although is perhaps most well known in machine learning for use in unsupervised learning problems, such as density estimation and clustering.
The most discussed application of the EM algorithm is for clustering with a mixture model.
In the EM algorithm, the estimation-step would estimate a value for the process latent variable for each data point, and the maximization step would optimize the parameters of the probability distributions in an attempt to best capture the density of the data. The process is repeated until a good set of latent values and a maximum likelihood is achieved that fits the data.

E-Step. Estimate the expected value for each latent variable.
M-Step. Optimize the parameters of the distribution using maximum likelihood.

Step 1: We have a set of missing or incomplete data and another set of starting parameters. We assume that observed data or the initial values of the parameters are generated from a specific model.
Step 2: Based on the observable value in the observable instances of the available data, we will predict or estimate the values in the unobservable instances of the data or the missing data. This is known as the Expectation step (E – step).
Step 3: Using the data generated from the E – step, we will update the parameters and complete the data

set. This is known as the Maximization step (M – step) which is used to update the hypothesis.

E step - calculate for each data point, the probability that datapoint belongs to cluster c.

M step - for each cluster c, calculate the fraction of points allocated to cluster c and update parameters.

The last step is to check if the values of latent variables are converging or not. If it gets "yes", then stop the process; else, repeat the process from step 2 until the convergence occurs.

Convergence is defined as the specific situation in probability based on intuition, e.g., if there are two random variables that have very less difference in their probability, then they are known as converged. In other words, whenever the values of given variables are matched with each other, it is called convergence.

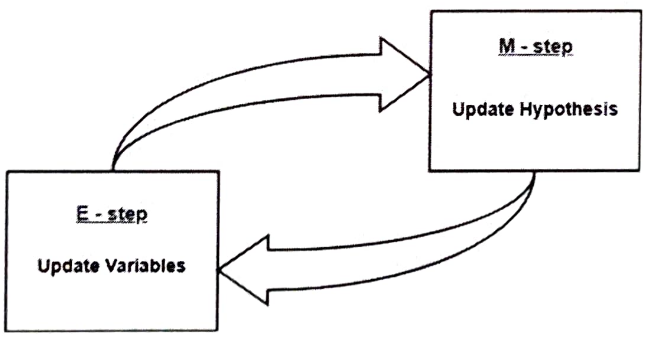
What Is The Maximum Likelihood Estimation?

In terms of statistics, maximum likelihood estimation is a method that helps to estimate all parameters of the probability distribution. The same works by maximizing a likelihood function for making probable the observed data for any statistical model.

However, the Maximum Likelihood mode comes with a big limitation. This is its assumption that data is complete as well as fully observed. It never mandates that a model could actually access all data. It goes on to assume that all variables that are relevant to a model are already present. The reality is that some relevant variables could remain hidden, leading to inconsistencies. Such hidden variables causing inconsistencies are termed Latent Variables.

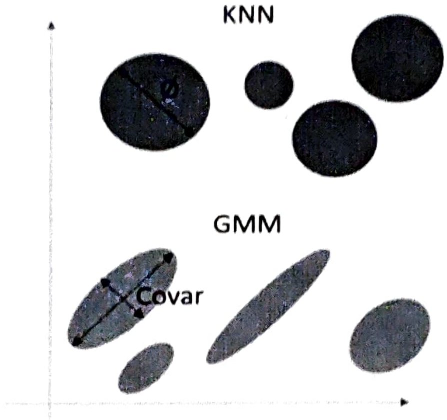
The Relevance of EM Algorithm

In the presence of a latent variable, the traditional maximum estimator won't work as you anticipate. Find an appropriate model parameter in the presence of a latent variable by employing the Expectation-Maximization or EM algorithm for machine learning.



A critical point for the understanding is that these gaussian shaped clusters must not be circular shaped as for instance in the KNN approach but can have all shapes a multivariate Gaussian distribution can take.

That is, a circle can only change in its diameter whilst a GMM model can (because of its covariance matrix) model all ellipsoid shapes as well. See the following illustration for an example in the two dimensional space.



Applications of the EM Algorithm:

- The latent variable model has plenty of real-world applications in machine learning.
- It is used in unsupervised data clustering and psychometric analysis.
- It is also used to compute the Gaussian density of a function.
- The EM algorithm finds extensive use in predicting the Hidden Markov Model (HMM) parameters and other mixed models.
- EM algorithm finds plenty of use in natural language processing (NLP), computer vision, and quantitative genetics.
- Other important applications of the EM algorithm include image reconstruction in the field of medicine and structural engineering.