Expectation-Maximization /cite 1 is a clustering algorithm which has an iterative approach. It basically computes maximum likelihood estimates iteratively and updates the distribution parameters (phi, mu, sigma) according to this likelihood information. This update can be implemented according to the log-likelihood and log-posterior. /cite2 /frank)

In this homework, we implement this solution to cluster a dataset which is a Gaussian Mixture model, includes 3 different Gaussian distribution.

We are trying to find best parameters theta (phi,sigma,mu) to maximize the log-likelihood of each data to appropriate distribution (here we can think distribution as clusters). /cite3 (moon) There is a paradox here. Since we don’t know the distributions from the beginning, we can not calculate the likelihood. Since we can not calculate the likelihood, we can not maximize it and find the appropriate Gaussian distributions of dataset.

To break this loop and start calculation, we determine random inital parameters for 3 Gaussian distributions: mu0,mu1,mu2,sigma0,sigma1,sigma2,phi0,phi1,phi2

1. Expectation

In this part, by using the parameters, we calculate the likelihoods of all the data points according to the different distributions. Than by using the formula down below, we calculate a posterior probability (sometimes people refer this responsibility) /cite4(gebru) for each of the data points and for each of the distributions. We will use this probabilities to update the parameters in “Maximization” step.

1. Maximization

By using the probabilities come from the expectation step, we update the phi, mean and sigma parameters in maximization step, to increase the each point’s likelihood to appropriate Gaussian distribution. Which also means to devide data points into clusters in most correct way.