**Summer Semester 2020 SNLP Assignment 9**

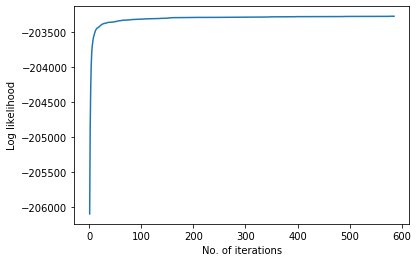
**Name: Awantee Deshpande  
Id: 2581348  
Email:** [**s8awdesh@stud.uni-saarland.de**](mailto:s8awdesh@stud.uni-saarland.de)

**Name: Lakshmi Rajendra Bashyam  
Id: 2581455  
Email:** [**s8laraje@stud.uni-saarland.de**](mailto:s8awdesh@stud.uni-saarland.de)

1) Some kinds of data result from drawing from multiple distributions. If the parameters of those distributions are known, it is simple to determine which distribution a data point comes from. But if those parameters are not known, expectation maximization can be used to determine the likelihood of a data point belonging to a particular distribution using some latent variable.

EM is particularly used to fit a model in an unsupervised setting. One well known use-case of Expectation Maximization is in the Baum-Welch algorithm for Hidden Markov Models. It is used to calculate the state transition probabilities and the emission probabilities to maximise the given data observation.

2) The code for the expectation and maximisations steps is in the functions E\_step and M\_step respectively in em\_wsd.py  
The change in log likelihood is modeled below as follows



3) The outputted ordered frequencies per cluster are

ordered list of senses within cluster 0: [('HARD1', 918), ('HARD2', 223), ('HARD3', 150)]

ordered list of senses within cluster 1: [('HARD1', 379), ('HARD2', 136), ('HARD3', 108)]

ordered list of senses within cluster 2: [('HARD1', 2158), ('HARD2', 143), ('HARD3', 118)]

Time: 53.60577321052551

4) EM does not find the global optimum. Similar to gradient descent, it only converges to a point where the log likelihood ~ 0 w..r.t the parameters. The search is local and and largely dependent on the first initialisations of the parameters.