**Website Recommendation Machine**

**Implemented by Bayesian Clustering Algorithm**

**Input:**

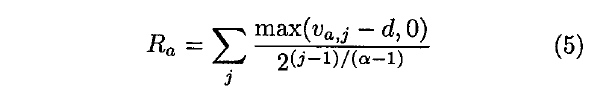
**A sparse matrix, columns are the number of website, rows are the users. If a user visited a website, the value is 1 else 0.**

**Output:**

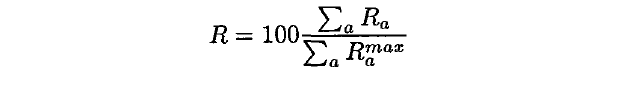
**A recommended list of websites for every user**

**Evaluation Criteria: Scoring Rank**

**For a fixed user:**



**Averaging over all the user:**

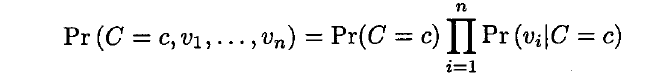


**The denominator is the summation of maximum achievable score for a user.**

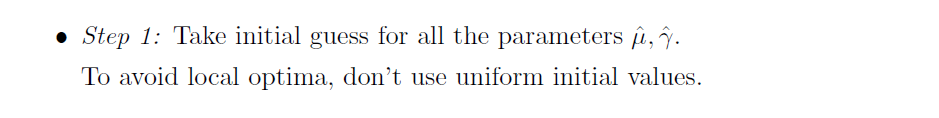
**Result: Around 47% -50% (on the test data)**

**Bayesian clustering Model intuition:**

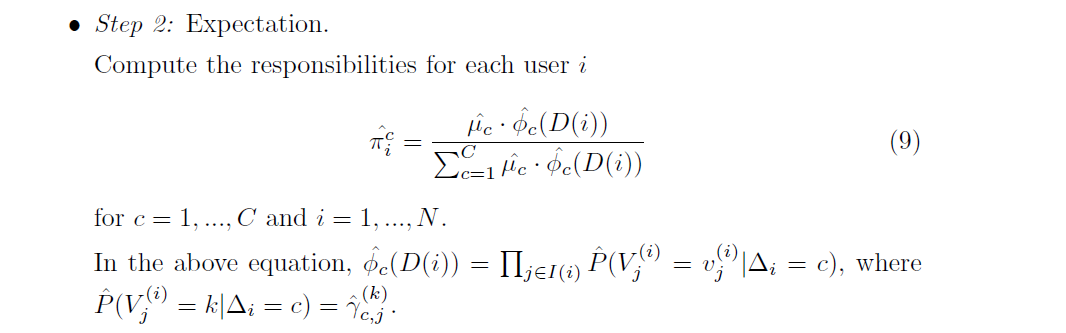
**Based on the user’s past visiting history, we cluster similar users into K different groups. Given a group i, the event of a user will click a website is binominal. This is a type of naïve Bayesian clustering, all the click events are conditionally independent given a group.**



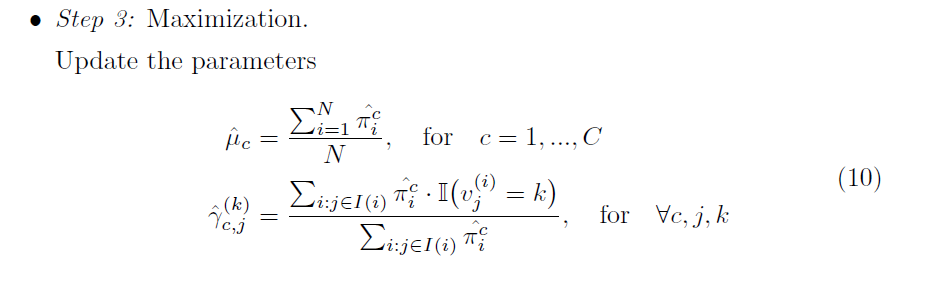
**We implement EM algorithm for estimating the parameter.**



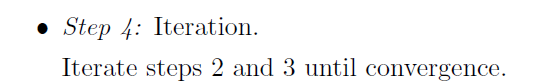
**We draw samples from Dirichlet distribution for initial value of the parameter.**

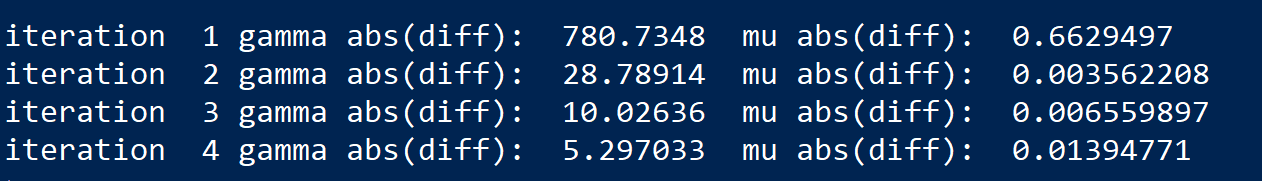


**We found the magnitude is well handled by the R. In our case, we didn’t do any particular logarithm process.**



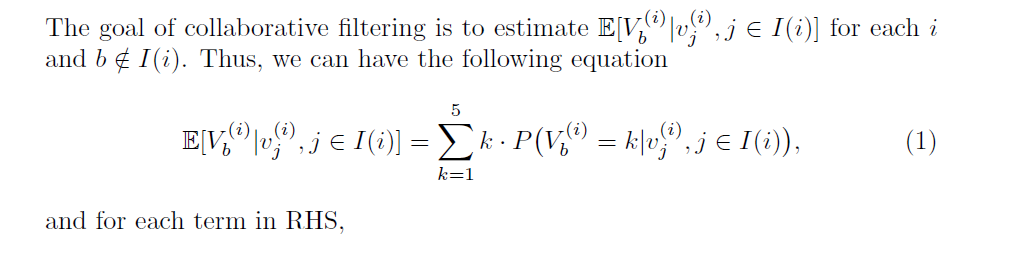
**These are the MLE estimator from the likelihood function. Since our model are exponential family, they are in closed form.**





**We choose the threshold to be sum(|gamma(i)-gamma(i+1)|)<10.**

**Predicted Scoring**



**We return a list movie, on that list all the movie are listed, but the order for each user is the decreasing order of the score.**

**Model Evaluation**

**In the end, we compare the rank score for each choice of K.**

**K=3 Rank Scoring 47.7**

**K=4 Rank Scoring 42.57**

**K=5 Rank Scoring 46.04**