Weekly Report III

Donglai Wei

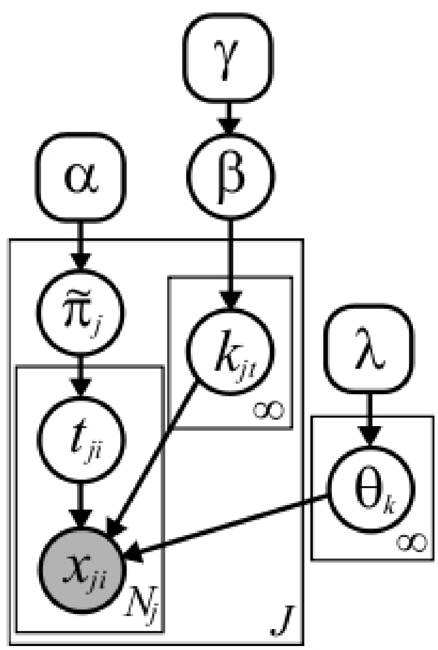
2010.5.4

1. HDP Gaussian Mixture Model

0) Notation:

Observations:
$\vec{x} = (x_{(11)}, ..., x_{(JN_j)})$ J Restaurants,N_j customers for each
Hidden Variables:
$\vec{z}$: assignments of table for customers($\vec{t}$) and dish for tables($\vec{k}$) from HDP;
θ: mean(μ) and covariance matrix(Ω) of Gaussian distributions;
Hyperparameters:
λ:(m_0, B_0, η_0, ξ_0 for μ, Ω and α, γ for $\vec{z}$)

1) Graphical Model for HDP:



2) Generative Model:
Likelihood Term:
$p(x_n, \theta|\lambda, z_n) = \mathcal{N}(x_n|z_n, \mu, \Omega)\mathcal{N}(\mu|m_0, \xi_0\Omega)\mathcal{W}(\Omega|\eta_0, B_0)$
Allocation Term:
$p(\vec{z}|\lambda) = \Pi_{j=1}^{J}[\frac{\Gamma(\alpha)}{\Gamma(n_{j..}+\alpha)}\Pi_{t=1}^{m_{j.}}(\Gamma(n_{jt.}))]\alpha^{\sum_{j=1}^{J} m_{j.}} \times \frac{\Gamma(\gamma)}{\Gamma(T+\gamma)}\Pi_{k=1}^{K}[\Gamma(m_{.k})]\gamma^K$

3) Marginal Probability given $\vec{z}$:
Margianlizing out θ,we get the negative of the log probability:

$log(p(x|z, \lambda))$:
$=$ (Likelihood)$- \sum_{k=1}^{K}[\frac{Dn_{..k}}{2}log\pi + \frac{D}{2}log\frac{\xi_k}{\xi_0} + \frac{\eta_k}{2}logdet(B_k) - \frac{\eta_0}{2}logdet(B_0) - log\frac{\Gamma_D(\frac{\eta_k}{2})}{\Gamma_D(\frac{\eta_0}{2})}]$
-
(Allocation:)$\underline{\sum_{j=1}^{J}\sum_{t=1}^{m_{j.}}[\frac{1}{m_{j.}}log\frac{\Gamma(n_{j..}+\alpha)}{\Gamma(\alpha)} - log(\Gamma(n_{jt.}) - log\alpha] + \sum_{k=1}^{K}[\frac{1}{K}log\frac{\Gamma(T+\gamma)}{\Gamma(\gamma)} - log(\Gamma(m_{.k}) - log\gamma]}$

2. Toy Data Set

0) Setting

(I) 9 Restaurants,200 customers for each,

(II) Generate the table assignment for each customer from a Dirichlet Process(concentration parameter α) in each restaurant.

(III) Generate the dish assignment for each table from a Dirichlet Process (concentration parameter γ)

(IV) Generate datas for each dish from one of the pre-defined 14 seperated 2-D Gaussian

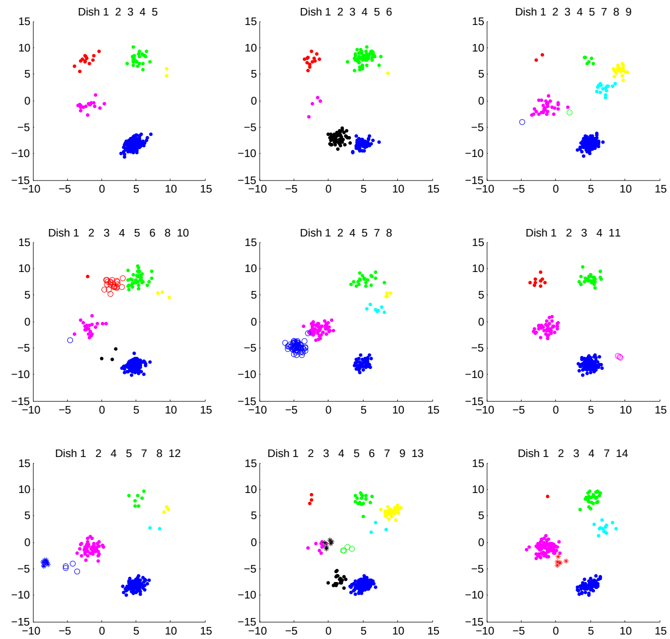


Figure 1: 9 Restaurants sharing same menu of dishes

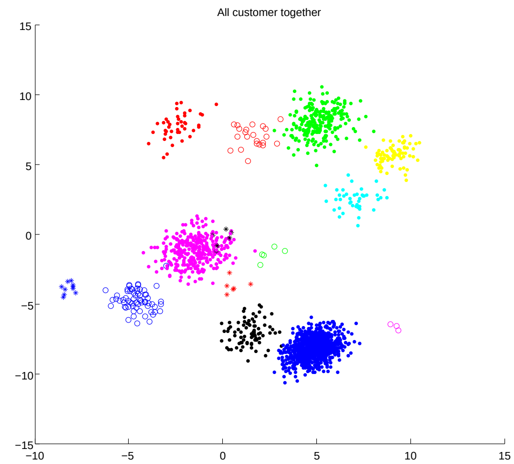


Figure 2: All customers together

1) Ground Truth

3. Search $\vec{z}$

The Goal is to find the $\vec{z}$ which gives the highest marginal probability P=$p(x|z,\lambda)$.

1) Local Search

Heuristically, we first try out the local search.

(I) Outline:

(a) Initialization: Every customer has his own table and every table has its own dish

(b) Iteration:

While there are still some local changes to increase P:

(i) In Random order, assign every customer the table in his restaurant which increases P mostly conditioning on other cutomers unchanged

(ii) In Random order, assign every table the dish which increases P mostly conditioning on other tables unchanged

end

(II) Result:
The clustering result for Restaurant 2,5,8 is not good enough, which have big clusters.

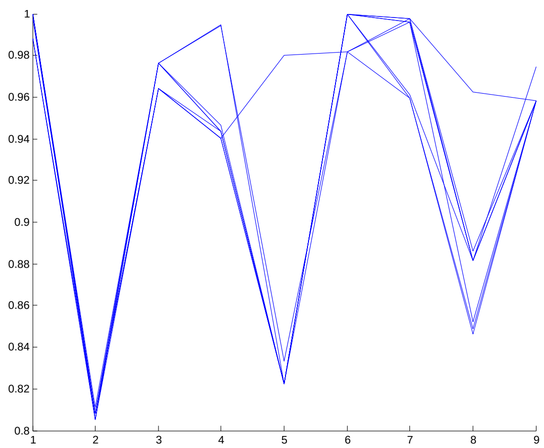


Figure 3: random index for each restaurant of one random search order

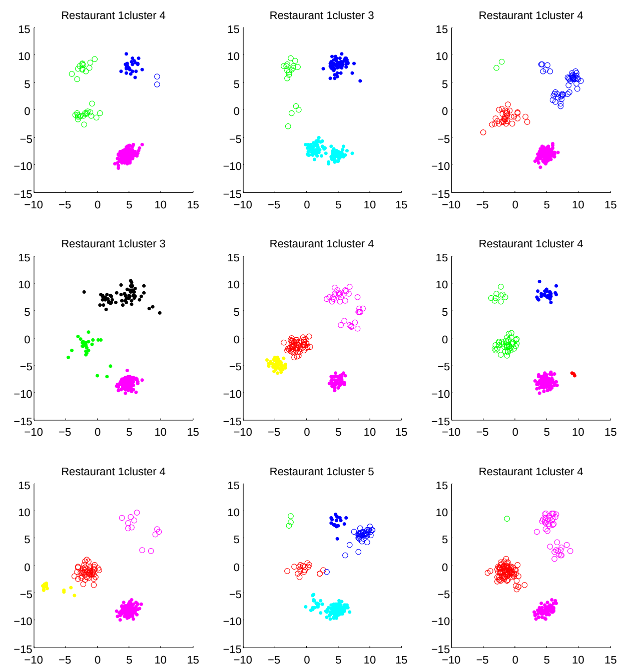


Figure 4: cluster result of one random local search order

2) Local Search+Split&Merge

After local search gets stuck at a local maxima, we try out split and merge(bigger moves) to find better maxima.

(I) Outline:

(a) Initialization: Start from the Fixed Point(local maxima) from Local Search

(b) Iteration:

While one of the last N moves of split or merge moves increases P:
Switch ceil(4*rand())

(i) case 1: In Random order, use k-means++ to split every dish into 2 dishes and accept it only if it increases P

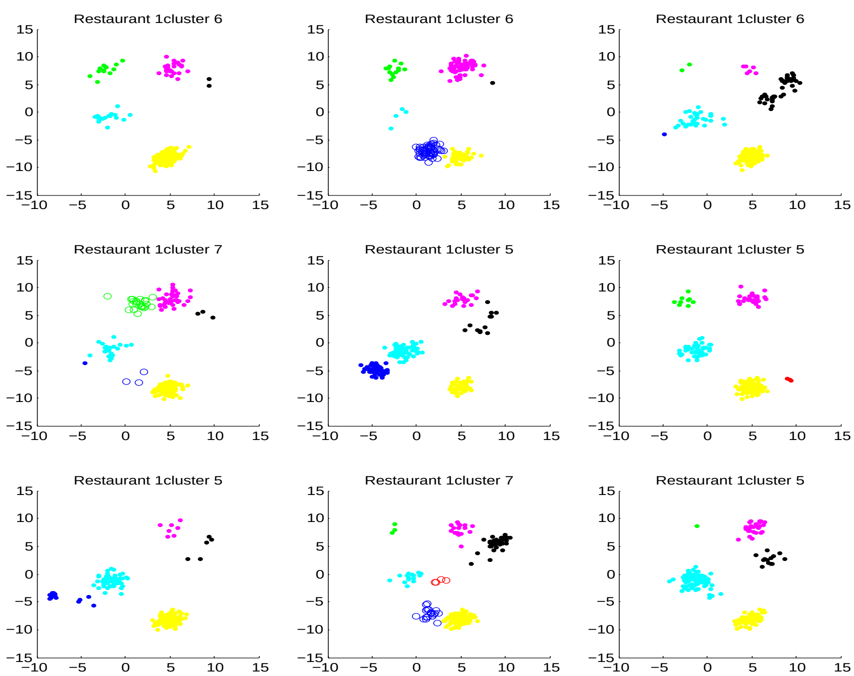
(ii) case 2: In Random order, use k-means++ to split every table into 2 tables and accept it only if it increases P

(iii) case 3: In Random order, merge every dish to another one which increases P mostly conditioning on other dishes unchanged and reject it if all merge moves decrease P

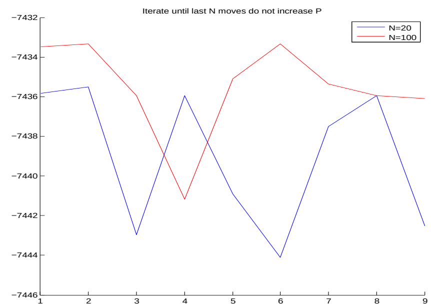
(iv) case 4: In Random order, merge every table to another one in the same restaurant which increases P mostly conditioning on other tables unchanged and reject it if all merge moves decrease P

end

(II) Result: Cluster result of one random split&merge order



[h]



For Split&Merge, we end the iteration until last N moves do not increase the log probability any more.We tried N=20 and N=100, which gives similar log probability.

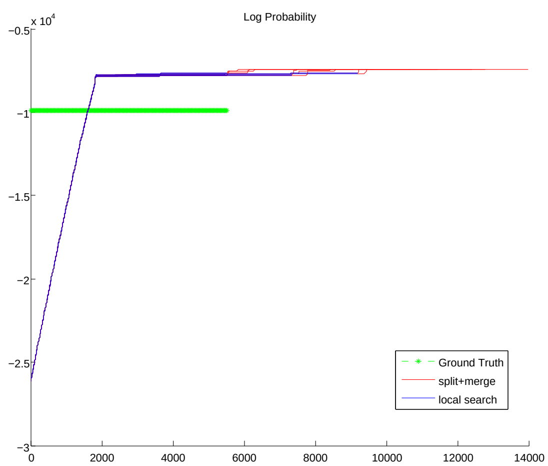


Figure 5: Log probability v.s. Steps

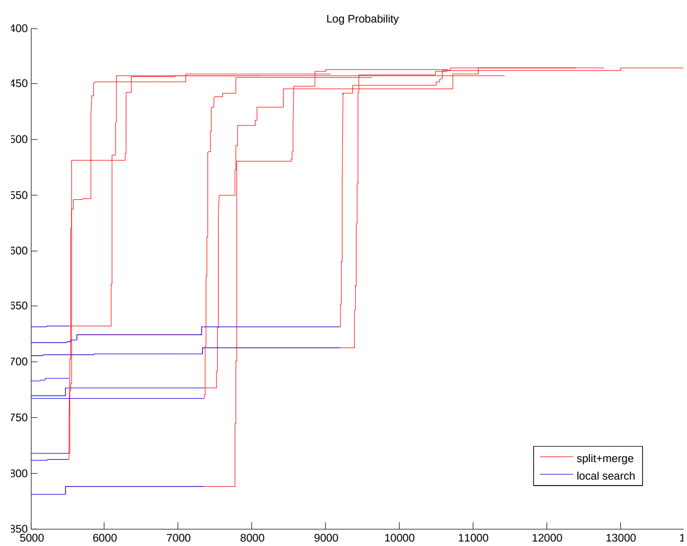


Figure 6: Closer LOOK of the convergence of log probability by split & merge

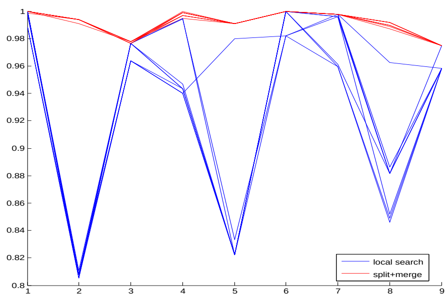


Figure 7: Rand Index Improvement Comparison

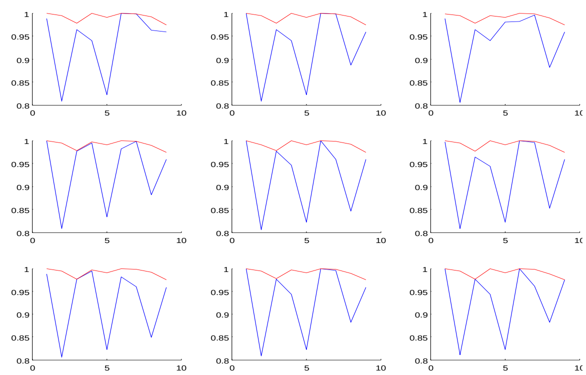


Figure 8: Rand Index Improvement in each restaurant