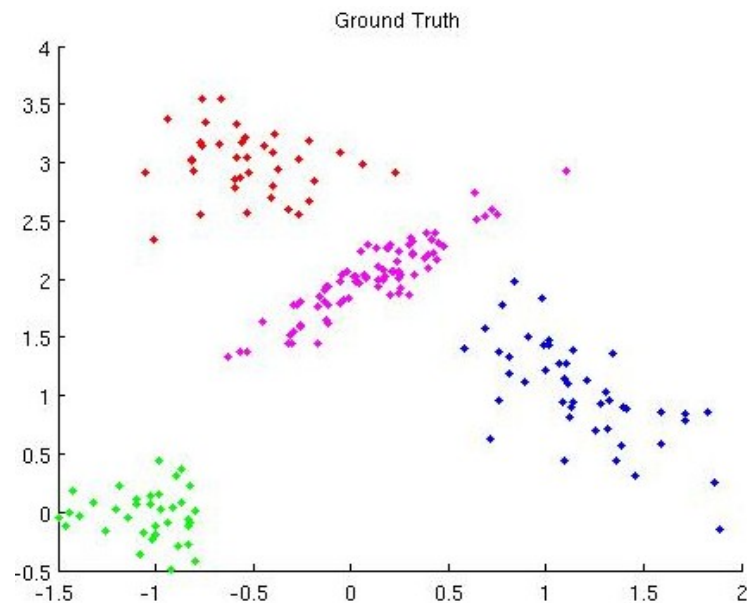# Comparison Among

Naïve Mean-field(bj), Collapsed Mean-field(csb) and ME algorithm
for Dirichlet Process Mixture Model

Donglai Wei
2010.3.18

0) Comparison of the Performance

Synthetic Data: 200 data points, Mixture of 4 Gaussian distributions.

Ground Truth

	Blei-Jordan (SIS 10 th)		Blei-Jordan (SIS 0th)		Collapsed (csb)		ME
	K=10	K=20	K=10	K=20	K=10	K=20	
#cluster (mean, std)	4.1 (0.3)	4.1 (0.3)	6.89 (1.27)	7.56 (1.70)	5.6 (1.6)	7.4 (2.0)	Bottom-up: 4 Top-down: 4 Local-search: a) top-down prior: 12 (merge into 5) a) bottom-up prior: 56 (merge into 12)
Rand Index	0.999 (1e-4)	0.999 (1e-3)	0.97 (0.02)	0.93 (0.04)	0.90 (0.11)	0.87 (0.09)	Bottom-up: 1 Top-down: 1 Local-search: a) top-down prior: 0.85 (0.97) b) bottom-up prior: 0.73 (0.82)

Algorithm1: Naïve Mean-Field (Blei-Jordan)
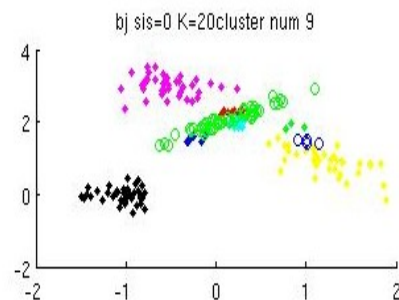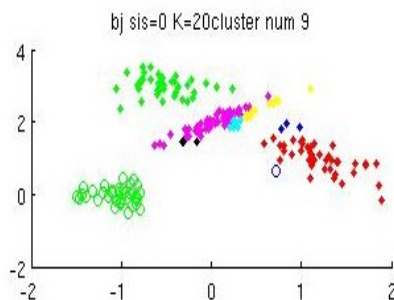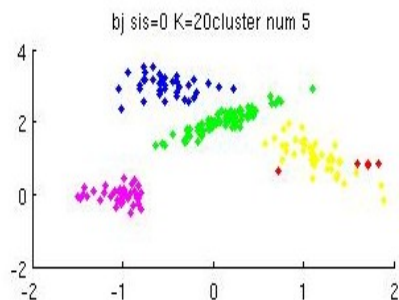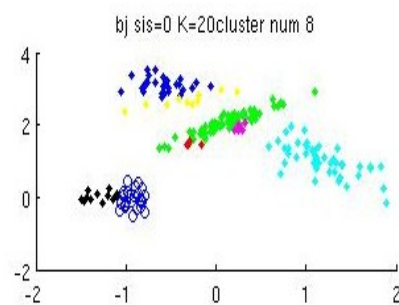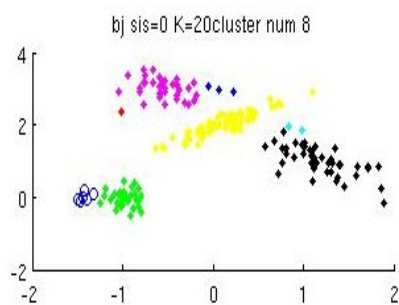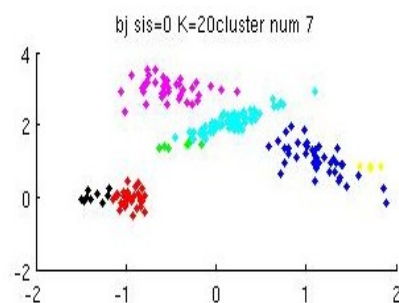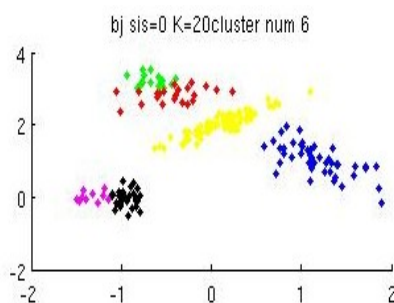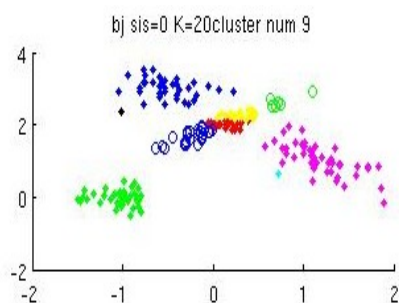
Test:
Initial Cluster number: 10, 20
Stopping threshold for free energy: 1e-10
Initialize the soft assignment with Sequential Importance Sampling : 0th, 10th

Remark:
1) Initialize with SIS 10th gives perfect result.
2) Without SIS, the performance of BJ algorithm may differ significantly even for our simple test case. Also some local minimas where the algorithm gets stuck are kind of funny, Like the one in the lower-left with 5 clusters, the red cluster is "splitted" by the yellow one.

	Blei-Jordan (SIS 10 th)		Blei-Jordan (SIS 0th)	
	K=10	K=20	K=10	K=20
#cluster (mean, std)	4.1 (0.3)	4.1 (0.3)	6.89 (1.27)	7.56 (1.70)
Rand Index	0.999(1e-4)	0.999 (1e-3)	0.97 (0.02)	0.93 (0.04)
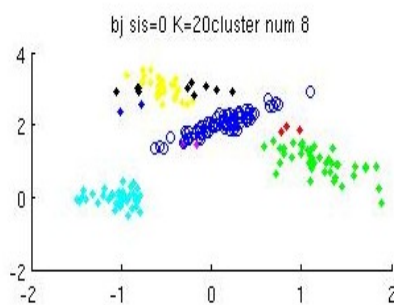
Algorithm2: Collapsed Mean-Field with Stick-Breaking Construction(csb)

Test:
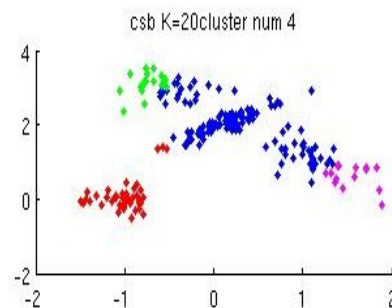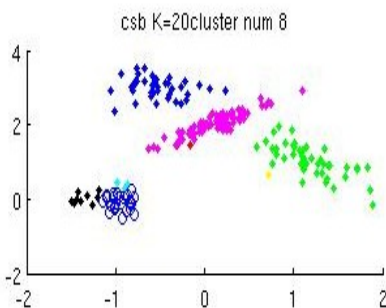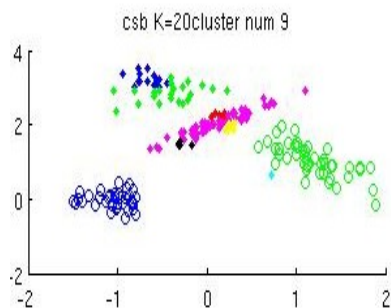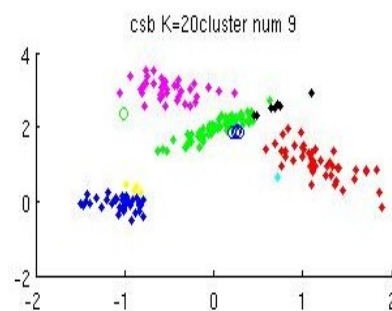Initial Cluster number: 10, 20
Stopping threshold for free energy: 1e-10

Remark:
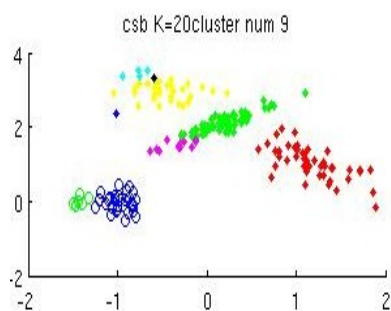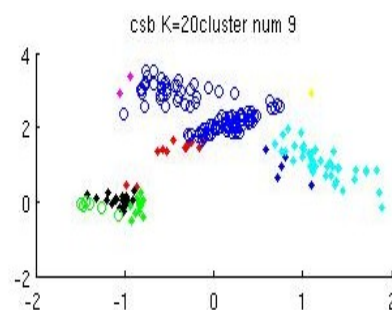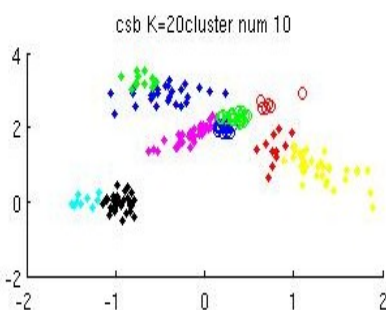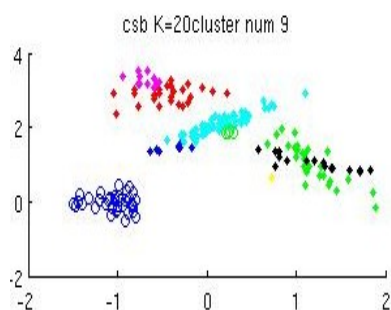a) Added with auxiliary parameters, collapsed methods may lead to more local minimas. Thus, different initiations of cluster numbers and soft cluster assignment vary significantly.
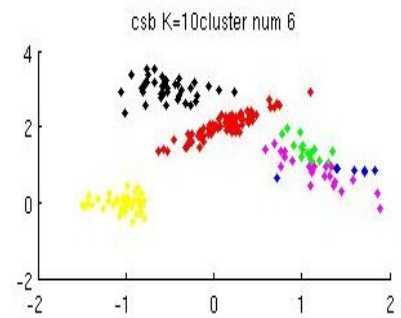
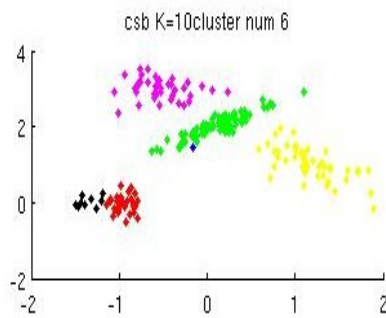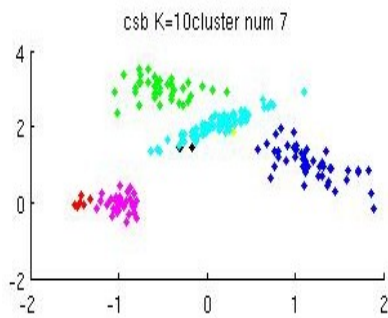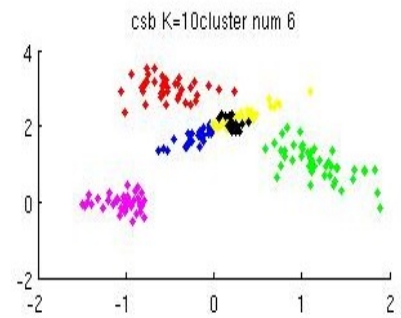|	Collapsed Meanfield (csb)	
	K=10	K=20
#cluster (mean, std)	5.6 (1.6)	7.4 (2.0)
Rand Index	0.90 (0.11)	0.87 (0.09)

K=20
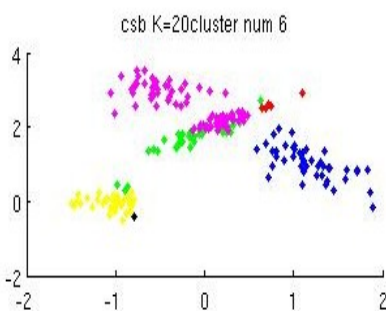
Algorithm2: Collapsed Mean-Field with Stick-Breaking Construction(csb)

(continued..) K=10
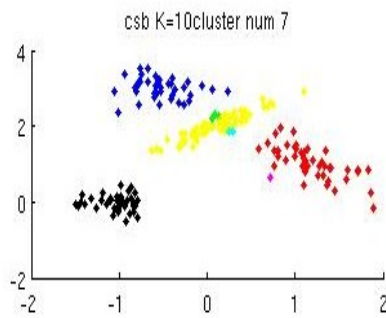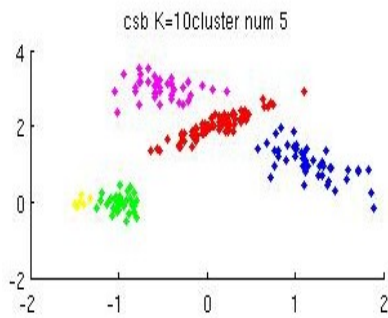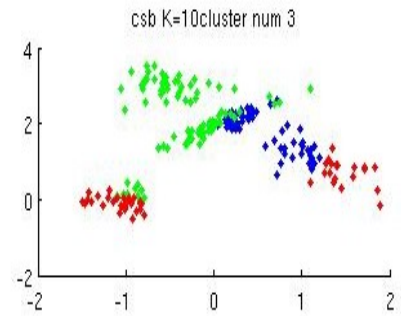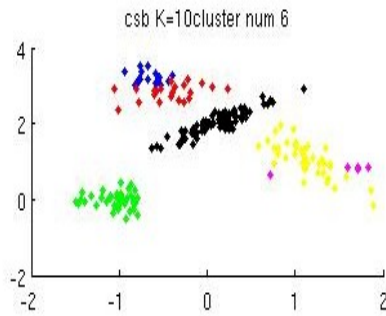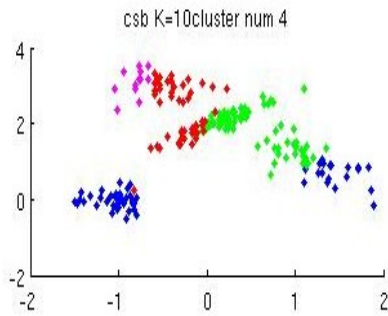
Algorithm 3: ME

Test:
Initial Cluster assignment:
1) Bottom-up, Local Search: 1:200 (every data forms its own cluster)
2) Top-down: 1 (every data belongs to the same cluster)
Prior for Local Search: Bottom-up, top-down

Remarks:
a) In this simple case, Bottom-up and Top-down method works perfectly well. But when facing with complicated real data, Bottom-up may be costly and Top-down may be heuristic.
b) For the local-search method, initialization of the hyper-parameter really matter. Claimed in the paper, the Bottom-up prior is uninformative since bottom-up method may be too greedy. But here informative may help

	ME
#cluster	Bottom-up: 4
	Top-down: 4
	Local-search:
	a) top-down prior: 12 (merge into 5)
	a) bottom-up prior: 56 (merge into 12)
Rand Index	Bottom-up: 1
	Top-down: 1
	Local-search:
	a) top-down prior: 0.85 (0.97)
	b) bottom-up prior: 0.73 (0.82)

ME-local search, top-down prior

ME-local search after merge, top-down prior

ME-local search first 20, bottom-up prior

ME-local search after merge, bottom-up prior