Experiments

October 2, 2013

1 Synthetic Dataset

`More mathematical formula`

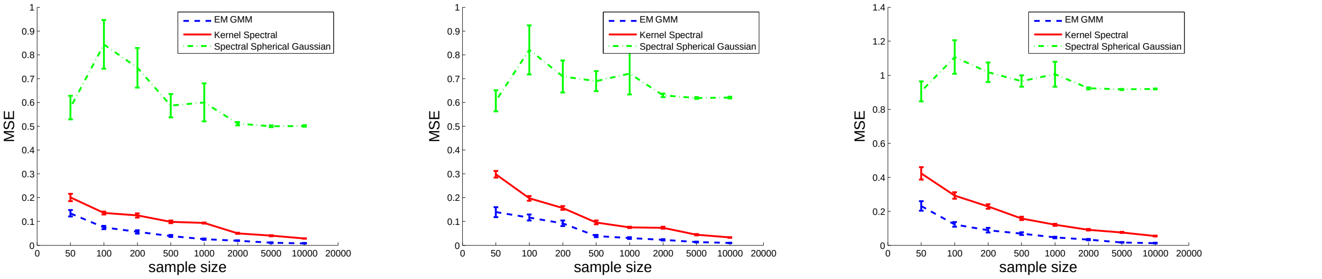
To verify the proposed algorithm, we compared with EM for mixture of Gaussians and spectral algorithm for mixture of spherical Gaussians (Hsu & Kakade, 2013). The assumption in (Hsu & Kakade, 2013) is very restricted, the means of the each Gaussian component should span a k-dimension subpsace, where k is number of components. Meanwhile, each component should be a spherical Gaussian. We also compared with another *nonparametric* spectral algorithms propsed by (Kasahara & Shimotsu, 2010) which uses histgram to approximate the conditional distribution. It could be thought as a special case of our method using delta kernel. It is predictable that their performance is not comparable to others because of the inferior histgram comparing to smooth kernel, the error is about 10 times to EM method. So that, we didn't plot it in the figures.

We generated synthetic data from the mixture models in four different settings, i.e., each view with different/same Gaussian/Gamma and Gaussian conditional distributions. In the Gamma/Gaussian mixture, we chose parameters to make the Gamma distribution more skew in latter two views while in the first view similar to Gaussian. In the Gamma/Gaussian mixture, we chose parameters to make the Gamma distribution more skew. For all the setting, we set the covariance of Gaussians be diagonal. The parameters for Gaussian and Gamma distributions are predefined to make sure they are not overlapped to much in the sense of the ratio of the distance between the mode of classes to the variance within the classes, i.e., for each pair of clusters, $\frac{(\mu_1-\mu_2)^2}{\sum_{i=1}^{1}\sigma_i^2}$ is big enough where $\mu = \frac{(\alpha-1)}{\beta}$ and $\sigma^2 = \frac{\alpha}{\beta^2}$ in Gamma distribution. We also varied the number of observations n, from 50 to $100,000$, and components k in range $[2, 3, 4, 8]$ in experiments to illustrate the convergence property of our algorithm. The mixture proportions are set to be $p_i = \frac{2i}{k(k+1)}, i \in \{1, \ldots, k\}$ which is highly unbalanced. It is worth to remark that when k becomes large but n is small, such setting is extremly difficult since only a little data will be generated from the first several clusters. For each n, k in each setting, we randomly sampled 10 sets of instances from the model for training.

For EM algorithm, since it is not guaranteed to get the global solutions each trial, we repeated it 10 times with random initialization and added regularization term to make sure the covariance parameters is valid. We selected the best kernel bandwidth for each view by evaluated on separated generated datasets. We measured the performance of algorithms by the $l2$-norm $||\sum_h p(x_i|h)p(h) - \sum_h \hat{p}(x_i|h)\hat{p}(h)||_2$ to the ground-truth marginal distribution of each view.

The results are plotted below. It is clear that both the nonparametric method and EM converge rapidly with the data increment in all experiments setting.

In mixture of Gaussian setting, the EM algorithm is best because the setting is fully satisfied to its assumption. While the spectral learning algorithm for sphereical Gaussian didn't perform

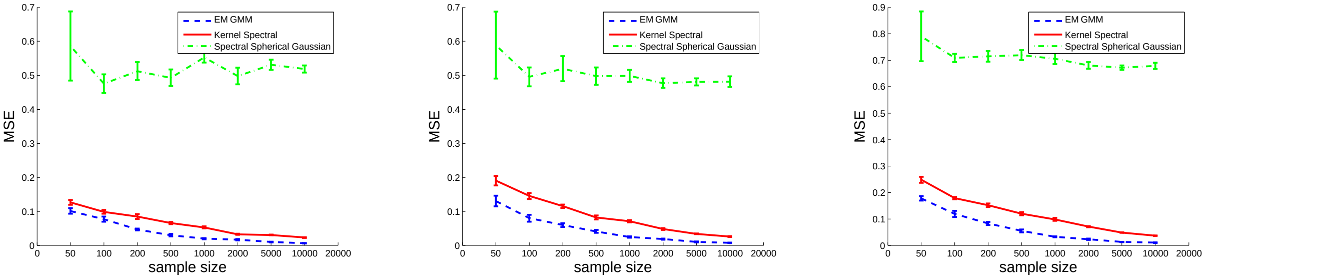


(a) View 1

(b) View 2

(c) View 3

Figure 1: The empirical results on synthetic dataset with different 2 Gaussian components in each view measured by the $l2$-norm between marginal distribution and ground-truth for each view.



(a) View 1

(b) View 2

(c) View 3

Figure 2: The empirical results on synthetic dataset with different 3 Gaussian components in each view measured by the $l2$-norm between marginal distribution and ground-truth for each view.

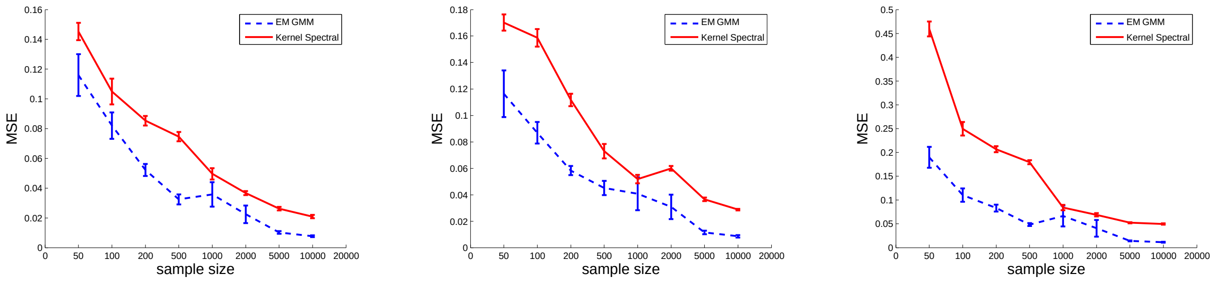
well since the assumption for the method is too restrict. It is reasonable that our nonparametric is a little bit worse than EM algorithm because of the complexity nonparametric form of probability distribution.

In the mixture of Gaussian/Gamma setting, the EM algorithm is good when data is less. However, our nonparametric spectral algorithm provided the best results when sample size is large enough. This phenomena demonstrated the trade-off between model complexity and sample size.

References

Hsu, Daniel and Kakade, Sham M. Learning mixtures of spherical gaussians: moment methods and spectral decompositions. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, ITCS '13, pp. 11–20, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1859-4.

Kasahara, Hiroyuki and Shimotsu, Katsumi. Nonparametric identification of multivariate mixtures. *Journal of the Royal Statistical Society - Series B*, 2010.

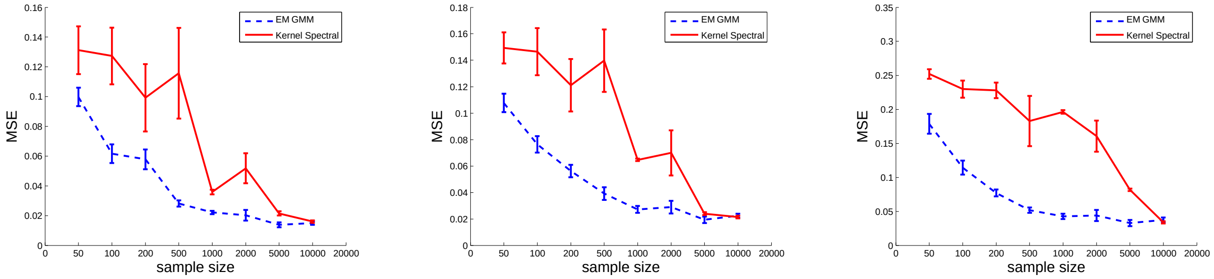


(a) View 1

(b) View 2

(c) View 3

Figure 3: The empirical results on synthetic dataset with different 4 Gaussian components in each view measured by the $l2$-norm between marginal distribution and ground-truth for each view.

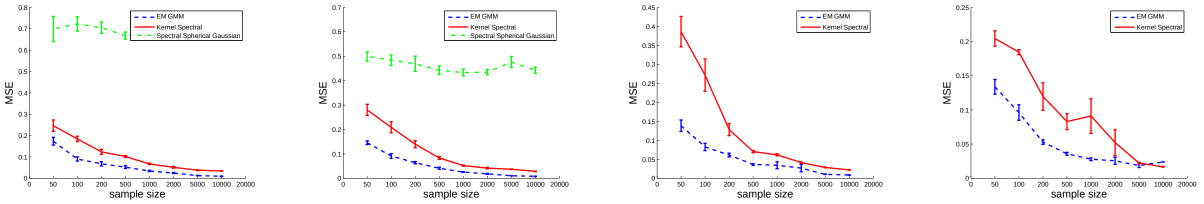


(a) View 1

(b) View 2

(c) View 3

Figure 4: The empirical results on synthetic dataset with different 8 Gaussian components in eavh view measured by the $l2$-norm between marginal distribution and ground-truth for each view.



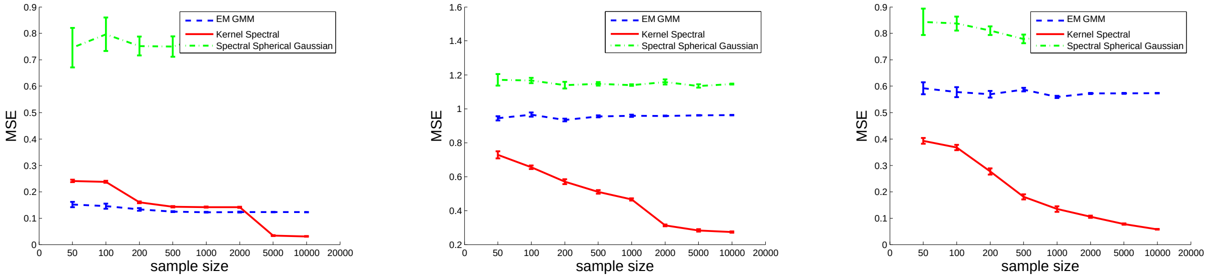
(a) 2 components

(b) 3 components

(c) 4 components

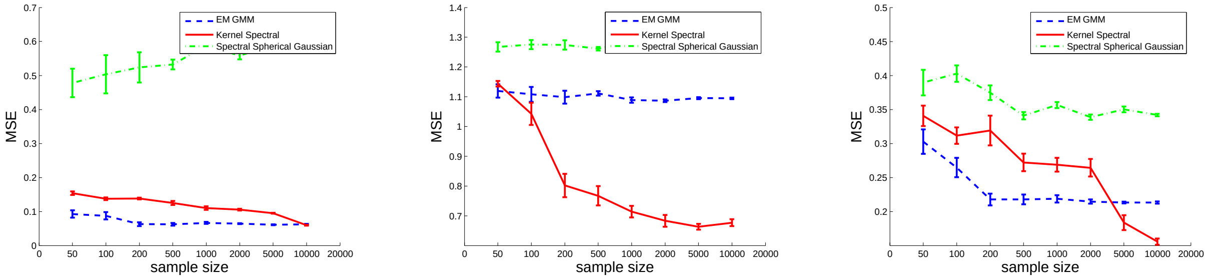
(d) 8 components

Figure 5: The empirical results on synthetic dataset with same Gaussian components in eavh view measured by the $l2$-norm between marginal distribution and ground-truth.



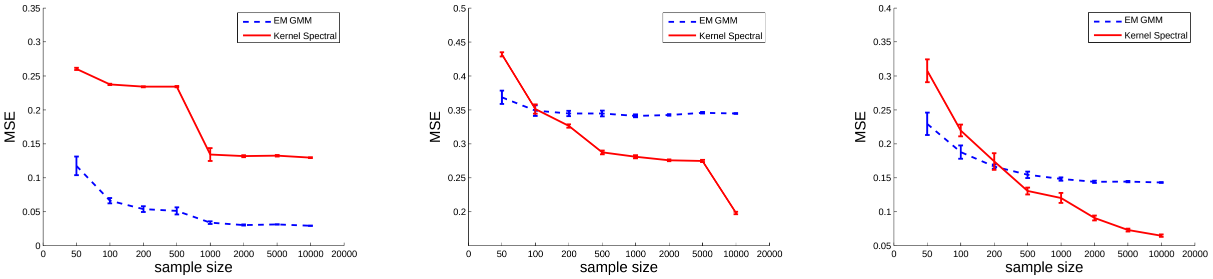
(a) View 1 (b) View 2 (c) View 3

Figure 6: The empirical results on synthetic dataset with different 2 Gaussian/Gamma components in eavh view measured by the $l2$-norm between marginal distribution and ground-truth for each view.



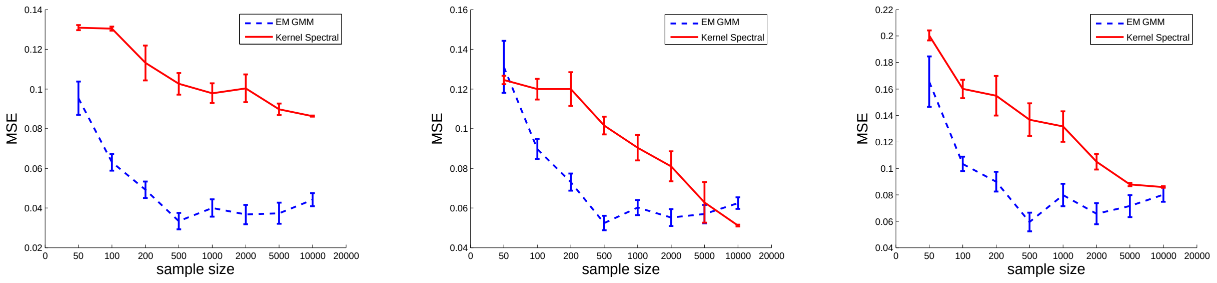
(a) View 1 (b) View 2 (c) View 3

Figure 7: The empirical results on synthetic dataset with different 3 Gaussian/Gamma components in eavh view measured by the $l2$-norm between marginal distribution and ground-truth for each view.



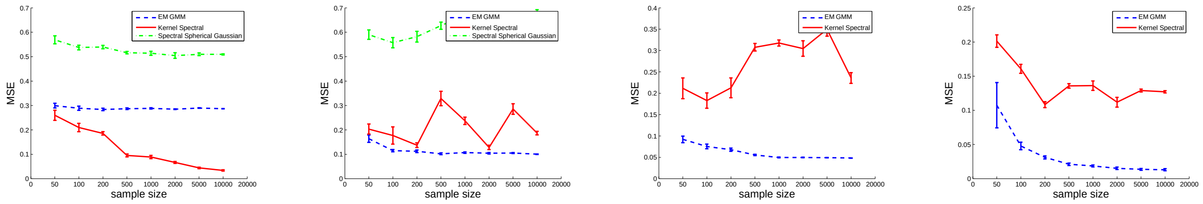
(a) View 1 (b) View 2 (c) View 3

Figure 8: The empirical results on synthetic dataset with different 4 Gaussian/Gamma components in eavh view measured by the $l2$-norm between marginal distribution and ground-truth for each view.



(a) View 1 (b) View 2 (c) View 3

Figure 9: The empirical results on synthetic dataset with different 8 Gaussian/Gamma components in eavh view measured by the $l2$-norm between marginal distribution and ground-truth for each view.



(a) 2 components (b) 3 components (c) 4 components (d) 8 components

Figure 10: The empirical results on synthetic dataset with same Gaussian/Gamma components in eavh view measured by the $l2$-norm between marginal distribution and ground-truth.