Robust Statistical Methods for Image Processing

Supervised Research Exposition (EE 451)

Kalpesh Patil (130040019)

Outline

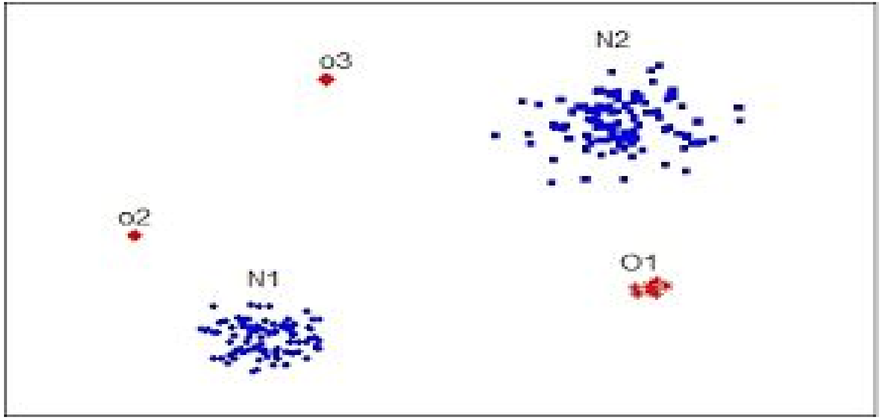
- Introduction
- Clustering Methods
  - Gaussian Mixture Models (GMM)
  - Student-t Distribution Mixture Models (SMM)
- Probabilistic Graphical Models
- Variational Bayesian Approximation
  - VBGMM
  - VBSMM
- Experiments and Results
  - Qualitative Comparison (GMM vs SMM)
  - Quantitative Comparison (GMM vs SMM)
  - Likelihood Comparison (VBGMM vs VBSMM)
- Applications: Functional MRI (fMRI)
- Conclusion and Future Work
- References

Outline

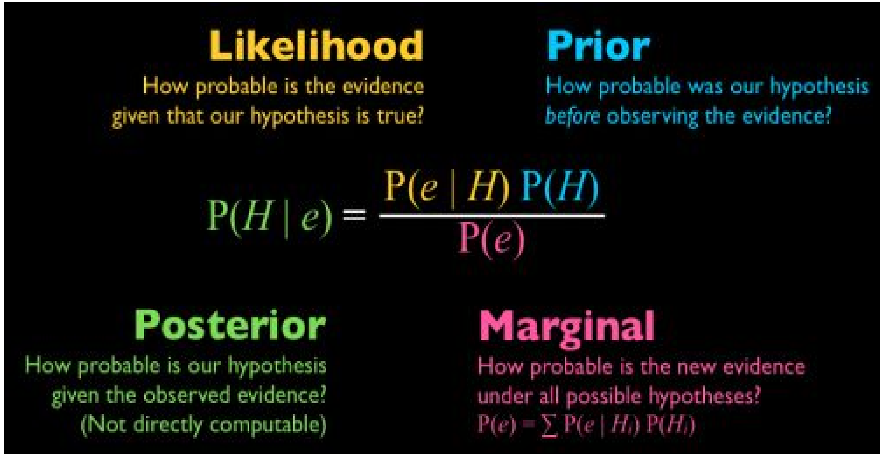
- **Introduction**
- Clustering Methods
    - Gaussian Mixture Models (GMM)
    - Student-t Distribution Mixture Models (SMM)
- Probabilistic Graphical Models
- Variational Bayesian Approximation
    - VBGMM
    - VBSMM
- Experiments and Results
    - Qualitative Comparison (GMM vs SMM)
    - Quantitative Comparison (GMM vs SMM)
    - Likelihood Comparison (VBGMM vs VBSMM)
- Applications: Functional MRI (fMRI)
- Conclusion and Future Work
- References

Introduction

- Robust Statistical Models
  - Presence of large outliers in practical data
  - Susceptibility of unsupervised clustering Techniques to such outliers



- Inferencing using statistical models
  - Bayesian estimation of parameters
  - Intractable integrals
  - Approximate methods



Outline

- Introduction
- **Clustering Methods**
  - Gaussian Mixture Models (GMM)
  - Student-t Distribution Mixture Models (SMM)
- Probabilistic Graphical Models
- Variational Bayesian Approximation
  - VBGMM
  - VBSMM
- Experiments and Results
  - Qualitative Comparison (GMM vs SMM)
  - Quantitative Comparison (GMM vs SMM)
  - Likelihood Comparison (VBGMM vs VBSMM)
- Applications: Functional MRI (fMRI)
- Conclusion and Future Work
- References

GMM

- Weighted mixture of Gaussian densities
- Means, covariance matrices and membership weights

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

- EM algorithm for parameter estimation
  - E step

$$\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$$

  - M step

$$\boldsymbol{\theta}^{\text{new}} = \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$$

Limitations of GMM

- Assumption that data-points belong to Gaussian distribution

- Fails to model large outliers
  Very less density on points away from centres

- Need of “heavy-tailed” distribution to model large outliers

Student-t Distribution

- Combination of infinitely many "scaled covariance" Gaussians

$$\int \phi(y_j; \mu, \frac{\Sigma}{u}) du$$

- Scale parameter (u) distributed according to Gamma density

- Functional form of t-distribution

$$f(y_j; \mu, \Sigma, \nu) = \frac{\Gamma(\frac{\nu+p}{2})}{|\Sigma|^{1/2}(\pi\nu)^{\frac{p}{2}}\Gamma(\frac{\nu}{2})}(1 + \delta(y_j; \mu, \Sigma)/\nu))^{-\frac{\nu+p}{2}}$$

$$\delta(y_j; \mu, \Sigma) = (y_j - \mu)^T \Sigma^{-1}(y_j - \mu)$$

SMM

- Low degrees of freedom implies heavier tails and thus robustness
- Modelling of data points in SMM

$$Y_j|(u_j, z_{ij} = 1) = N(\mu_i, \frac{\Sigma_i}{u_j})$$

- EM algorithm for parameter updates

$$U_j|(z_{ij} = 1) = gamma(\frac{v_i}{2}, \frac{v_i}{2})$$

More complex update equation than Gaussian
Doesn’t have closed form update for DoF

Outline

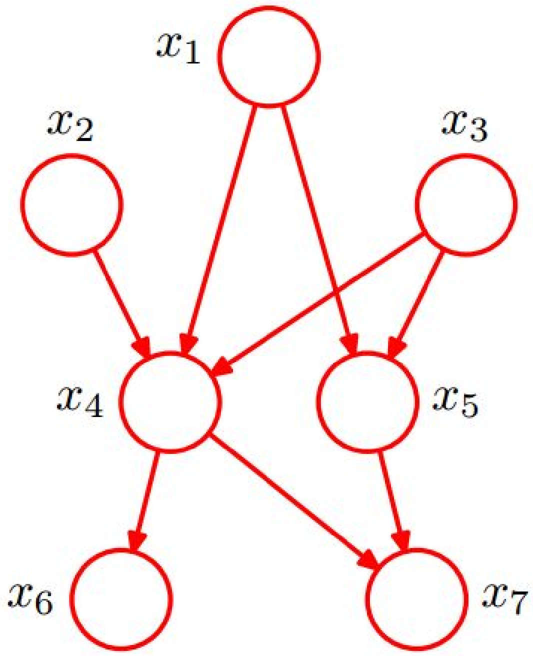
- Introduction
- Clustering Methods
  - Gaussian Mixture Models (GMM)
  - Student-t Distribution Mixture Models (SMM)
- **Probabilistic Graphical Models**
- Variational Bayesian Approximation
  - VBGMM
  - VBSMM
- Experiments and Results
  - Qualitative Comparison (GMM vs SMM)
  - Quantitative Comparison (GMM vs SMM)
  - Likelihood Comparison (VBGMM vs VBSMM)
- Applications: Functional MRI (fMRI)
- Conclusion and Future Work
- References

Probabilistic Graphical Models

- Used to denote "dependence" relationship amongst random variables

- Joint distribution using DAG

- Nodes: random variables
  Edges: Conditional dependence



$$p(x_1)p(x_2)p(x_3)p(x_4|x_1,x_2,x_3)p(x_5|x_1,x_3)p(x_6|x_4)p(x_7|x_4,x_5)$$

Outline

- Introduction
- Clustering Methods
  - Gaussian Mixture Models (GMM)
  - Student-t Distribution Mixture Models (SMM)
- Probabilistic Graphical Models
- **Variational Bayesian Approximation**
  - VBGMM
  - VBSMM
- Experiments and Results
  - Qualitative Comparison (GMM vs SMM)
  - Quantitative Comparison (GMM vs SMM)
  - Likelihood Comparison (VBGMM vs VBSMM)
- Applications: Functional MRI (fMRI)
- Conclusion and Future Work
- References

Variational Bayesian Approximation

Total evidence of data given model structure (independent of parameters)

$$Pr(X|H_M) = \int_\theta Pr(X|\theta, H_M)Pr(\theta|H_M)d\theta$$

Intractable integrals, hence approximations used

1. Numerical Approximation:
   Numerically solve integrals. e.g. MCMC method
2. Variational Bayesian Approach:
   Assumption: Joint distribution factorizes into components

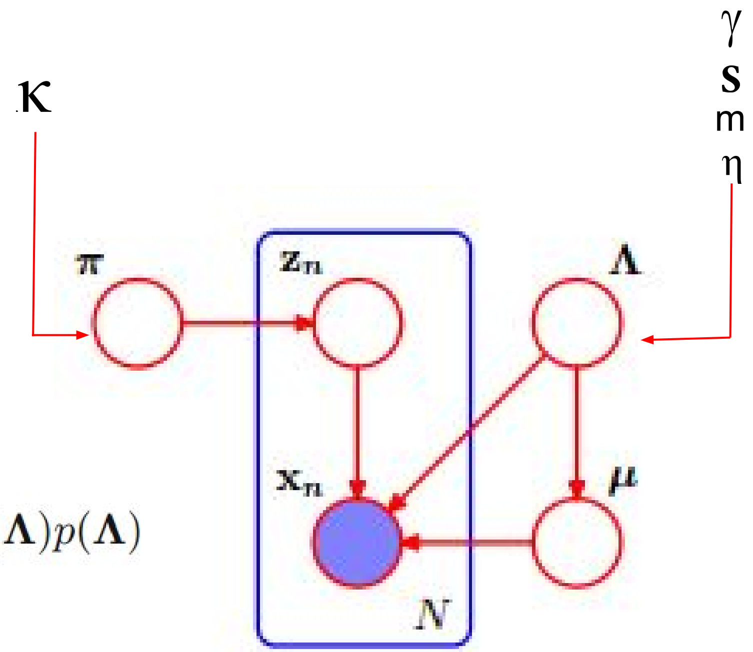
VBGMM

- Priors over parameters of model
  - Mean and (inverse) covariance:
    Gaussian-Wishart prior
  - Mixture proportions:
    Dirichlet distribution

- Joint probability

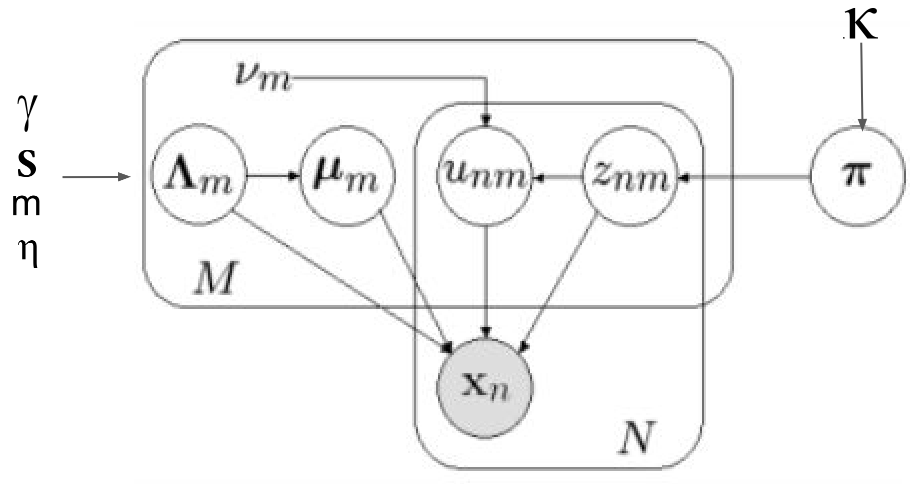
$$p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = p(\mathbf{X}|\mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\Lambda})p(\mathbf{Z}|\boldsymbol{\pi})p(\boldsymbol{\pi})p(\boldsymbol{\mu}|\boldsymbol{\Lambda})p(\boldsymbol{\Lambda})$$

- VBEM algorithm
  Variational Bayesian EM algorithm for optimization



VBSMM

- Priors over parameters of model
  - Mean and (inverse) covariance:
    Gaussian-Wishart prior
  - Mixture proportions
    Dirichlet distribution
  - Degree of freedom
    No prior



- Joint probability

$$Pr(\theta_s|H_M) = D(\pi|\kappa_0) \prod_{m=1}^{M} NW(\mu_m, \Lambda_m|\theta_{NW_0}).$$

- VBEM algorithm

Outline

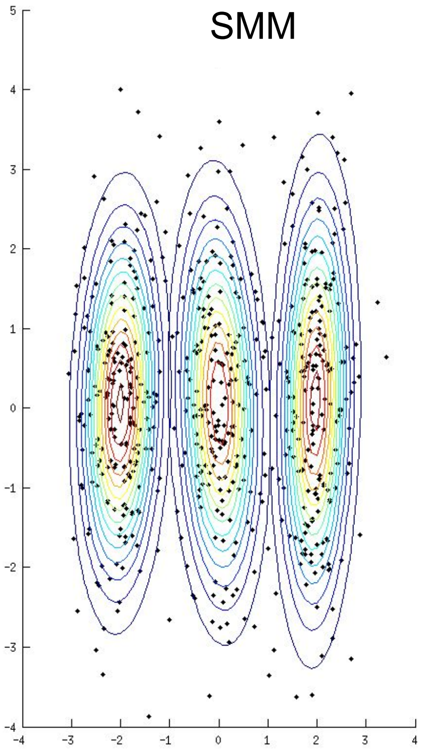
- Introduction
- Clustering Methods
  - Gaussian Mixture Models (GMM)
  - Student-t Distribution Mixture Models (SMM)
- Probabilistic Graphical Models
- Variational Bayesian Approximation
  - VBGMM
  - VBSMM
- **Experiments and Results**
  - Qualitative Comparison (GMM vs SMM)
  - Quantitative Comparison (GMM vs SMM)
  - Likelihood Comparison (VBGMM vs VBSMM)
- Applications: Functional MRI (fMRI)
- Conclusion and Future Work
- References

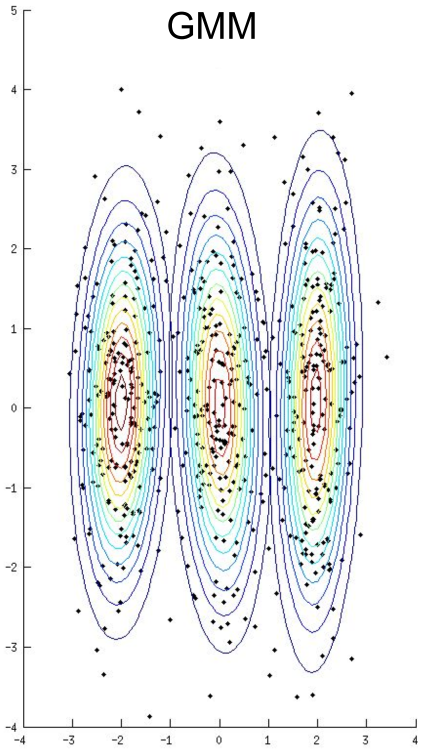
Experimental Setup

- 2D toy dataset consisting of 3 Gaussian components with common diagonal covariance matrix
- 200 data-points sampled from each component
- Outliers away from the cluster centres are added to the original data
- Try to fit GMM, SMM, VBGMM, VBSMM on data with and without outliers
- Also compute total likelihood for VB methods

GMM vs. SMM without Outliers

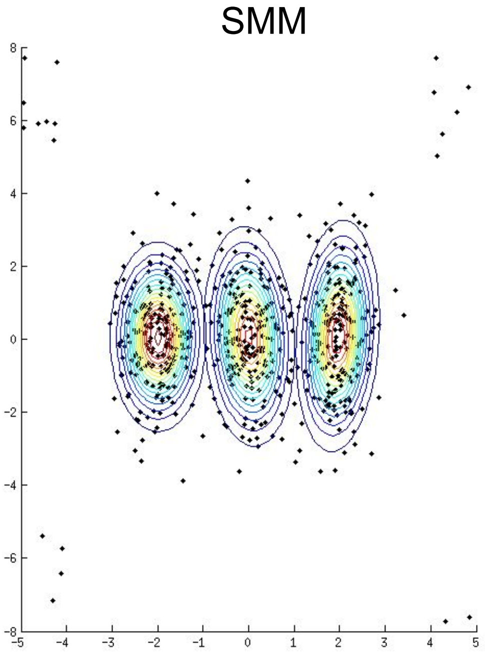
- Both GMM and SMM Perform almost equally good in absence of outliers

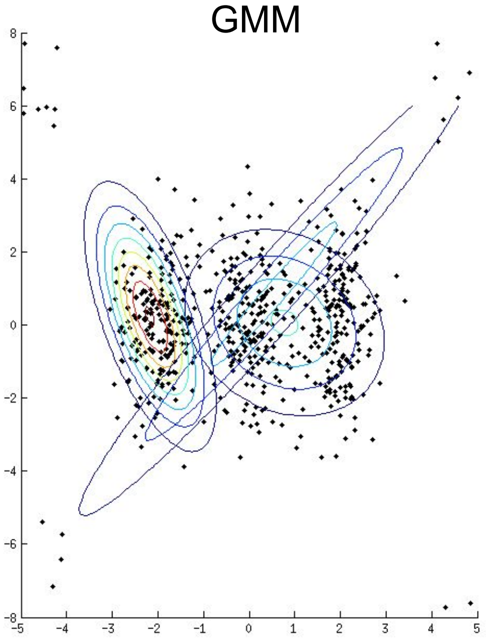




GMM vs. SMM with Outliers

- 20 outliers added to 600 datapoints
- GMM breaks down and performs poorly
- SMM shows robustness even in presence of outliers





Quantitative Analysis

- Average Euclidean distance between cluster centers

$$d_{cluster-centers} = \frac{1}{K}\sum_{k=0}^{K} ||c_k^{est} - c_k^{orig}||_2$$

- Mean cosine similarity between eigenvectors of covariance matrices

$$s_{eigen-cos} = \frac{1}{K}\sum_{k=0}^{K}\sum_{j=0}^{j=d} abs\left(\frac{v_{jk}^{est}.v_{jk}^{orig}}{||v_{jk}^{est}||_2||v_{jk}^{orig}||_2}\right)$$

Quantitative Analysis

- Results of GMM and SMM with and without outliers

	without outliers		with outliers	
Quantity	GMM	SMM	GMM	SMM
$d_{cluster-centers}$	0.0820	0.0782	1.1148	0.0883
$s_{eigen-cos}$	1.9987	1.9987	1.7896	1.9987

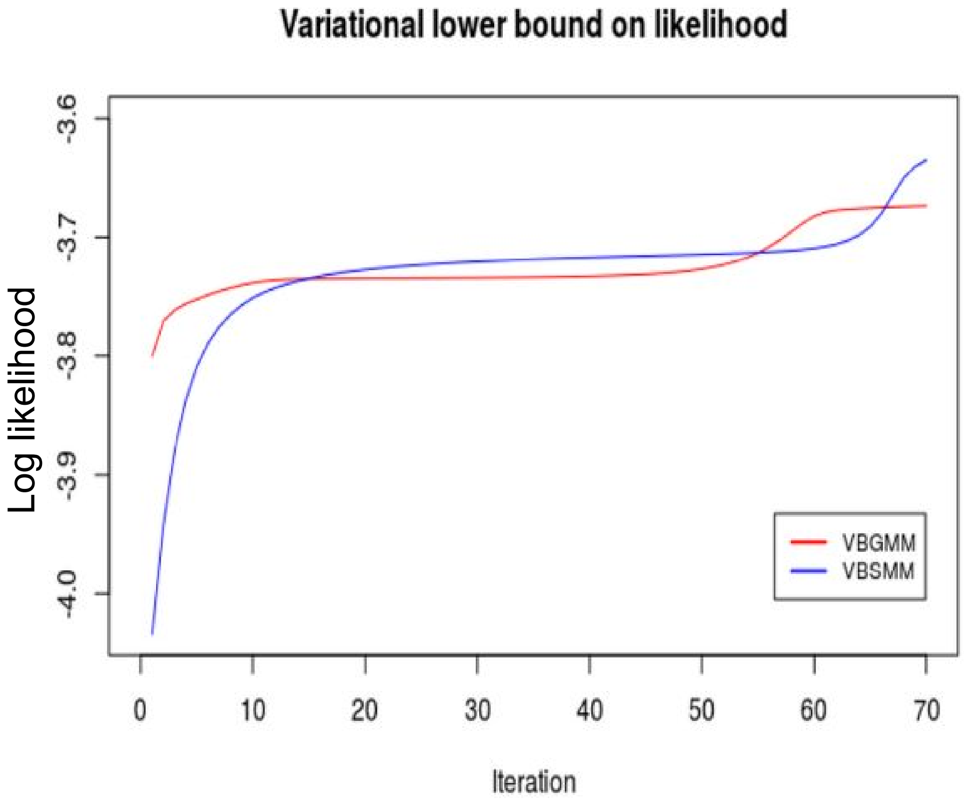
- Both perform equally well in the absence of outliers

- SMM clearly outperforms GMM in presence of outliers in both measures

VBGMM vs. VBSMM (without Outliers)

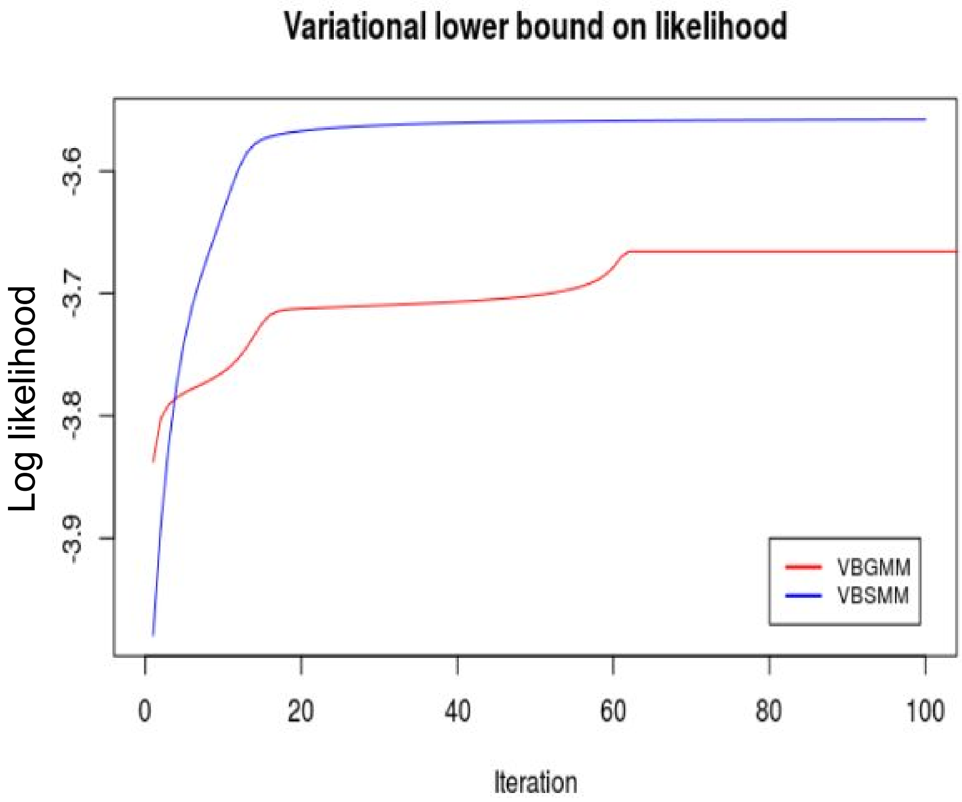
- Recall: Variational Bayesian approach allows to compute Complete data likelihood Independent of parameters

- In absence of outliers as EM iterations proceed, both VBGMM and VBSMM seem to converge to the same to the same value



VBGMM vs. VBSMM (with outliers)

- In presence of outliers VBSMM converge to a higher likelihood value than VBGMM
- VBSMM is able to justify data containing outliers better than VBGMM

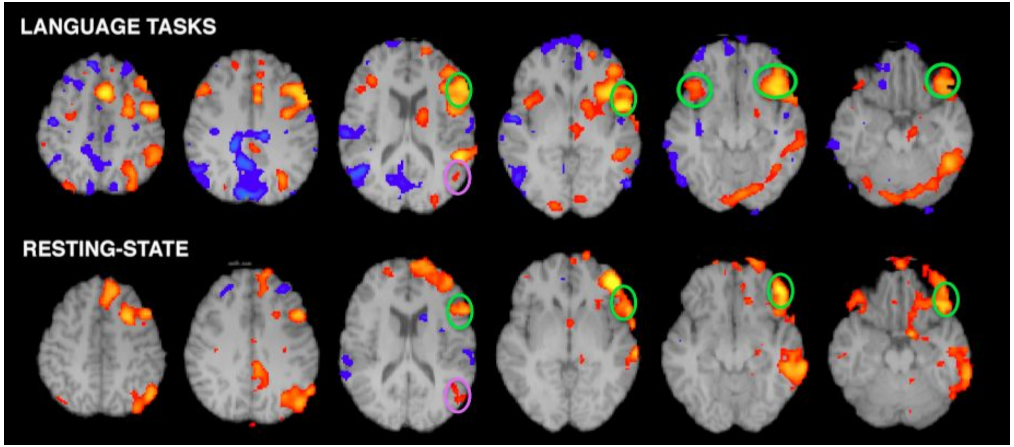


Outline

- Introduction
- Clustering Methods
  - Gaussian Mixture Models (GMM)
  - Student-t Distribution Mixture Models (SMM)
- Probabilistic Graphical Models
- Variational Bayesian Approximation
  - VBGMM
  - VBSMM
- Experiments and Results
  - Qualitative Comparison (GMM vs SMM)
  - Quantitative Comparison (GMM vs SMM)
  - Likelihood Comparison (VBGMM vs VBSMM)
- **Applications: Functional MRI (fMRI)**
- Conclusion and Future Work
- References

Functional-MRI (fMRI) Introduction

- Functional Magnetic Resonance Imaging
  4 dimensional object i.e. 3 dimensional voxel data collected at various time instants
- Analyzes temporal aspects of activations in various parts of the brain



fMRI Clustering

- Clustering of f-MRI time series data gives an idea about which parts of the brain are stimulated synchronously

- Contains large amount of outliers

- K-means, GMM, Ward's hierarchical clustering, Spectral clustering, c-means, ICA etc. have been studied before

- We believe that SMM with some regularization in the form of spatial smoothening prior would perform better due to its robustness against outliers in fMRI data

Outline

Conclusion and Future work

- SMM is more robust to large outliers than GMM
  Even in the Bayesian framework VBSMM performs better than VBGMM

- In future we would like to implement SMM or VBSMM on fMRI data

- Need to tackle curse of dimensionality while analyzing a very high dimensional data like fMRI.

- Modifications to incorporate spatial and temporal priors while performing clustering on actual fMRI data

References

[1] C. Bishop, “Pattern recognition and machine learning (information science and statistics), 1st edn. 2006. corr. 2nd printing edn,” Springer, New York, 2007

[2] G. J. McLachlan, S.-K. Ng, and R. Bean, “Robust cluster analysis via mixture models,” Austrian Journal of Statistics, vol. 35, no. 2, pp. 157– 174, 2006

[3] C. Archambeau and M. Verleysen, “Robust bayesian clustering,” Neural Networks, vol. 20, no. 1, pp. 129–138, 2007

[4] M. J. Beal, Variational algorithms for approximate Bayesian inference. University of London United Kingdom, 2003.

Thanks!

Questions?