Hierarchical Modal Clustering based on the Topography of Multivariate Mixtures

at

McGill University, 2006

Surajit Ray

sray@unc.edu, University of North Carolina at Chapel Hill
Introduction and univariate modality.

Multivariate modality.

Topography of Multivariate Mixtures.

Modal Clusters based on the topography of mixtures

Hierarchical Modal Clustering

Medical Imaging Application
Medical Image Segmentation

Axial (Transverse) CT of the Abdomen
Clustering Genes from gene expression Data

Data from Pugh et.al
Motivating Examples

Medical Image Segmentation
Clustering Genes from gene expression Data

Mixture of Distributions

Components vs Modes
Why study modality?

Univariate Mixtures

Multivariate Mixtures: Equal Variance
Multivariate Mixtures: Unequal Variance

Applications

Pi-plots and Analytical Solution
Modal Cluster
Hierarchical Modal Clustering with Example
Summary

---

**Mixture of Distributions**

A **$K$-component mixture** is given by

$$ f(x) = \sum_{j=1}^{K} \pi_j f_j(x; \lambda_j, \theta) \quad x \in \mathbb{R}^D $$

where $f_j(x; \lambda_j, \theta)$ are the **component densities** and $\pi_j$ are the **mixing proportions**. These are subject to the constraints:

$$ 0 \leq \pi_j \leq 1 \quad \forall j \quad \text{and} \quad \sum_{j=1}^{K} \pi_j = 1. $$

The parameter $\pi$ lives in a $K - 1$ simplex.
Mixture of Distributions

- Flexible way of modeling heterogeneous population.
- Data reduction technique: Appropriate Visualization.
- Directly applicable to model based clustering. [Ray 2003, Ray and Lindsay 2006]
Components vs Modes

5 component Bivariate Normal Mixture:

Motivating Examples
Medical Image Segmentation
Clustering Genes from gene expression Data
Mixture of Distributions
Components vs Modes
Why study modality?

Univariate Mixtures
Multivariate Mixtures: Equal Variance
Multivariate Mixtures: Unequal Variance

Applications
Pi-plots and Analytical Solution

Modal Cluster
Hierarchical Modal Clustering with Example

Summary

McGill University: Feb/March 2006 - slide #7
Components vs Modes

5 component Bivariate Normal Mixture: Only 3 modes
Why study modality?

- Potentially symptomatic to the underlying population structure.
- More **stable** and more **elemental** than components.

**Example:** To model a single t-distribution we may require mixture of “many” normals with same mean but different variance.
Why study modality?

- Potentially symptomatic to the underlying population structure.
- More **stable** and more **elemental** than components.

**Goal (1)**
- Study topography of **high dimension** normal mixtures in .
- Analytical and graphical solution for number of modes.

This research appeared in the October issue of the *Annals of Statistics*. 
Why study modality?

- Potentially symptomatic to the underlying population structure.
- More **stable** and more **elemental** than components.

**Goal (2)**
Use topography of normals for
- Dimension Reduction
- Merging component of mixtures. (Pruning)
- Developing **hierarchical modal clustering theory**.
For Univariate Normals: number of components $\geq$ number of modes.
Motivating Examples

Univariate Mixtures

Number of components ≠ Number of Modes

Multivariate Mixtures: Equal Variance

Multivariate Mixtures: Unequal Variance

Applications

Pi-plots and Analytical Solution

Modal Cluster

Hierarchical Modal Clustering with Example

Summary

---

**Number of components ≠ Number of Modes**

![Graphs showing bimodal and unimodal distributions](image)

**Bimodal**

Mixture of Normals 4 standard deviations apart

**Unimodal**

Mixture of Normals 2 standard deviations apart

**Conditions for bimodality of two univariate normals (Helguero, 1904)**

\[
X \sim \pi \mathcal{N}(\mu_1, \sigma) + (1 - \pi) \mathcal{N}(\mu_2, \sigma)
\]

\[
\left| \frac{\mu_2 - \mu_1}{\sigma} \right| \geq 2 \quad (\pi = \frac{1}{2})
\]

**Conditions for arbitrary univariate mixture in**

Modality: Equal Variance Bivariate Distribution

Bivariate normal with means (-1,-1) and (1,1)
Modality: Equal Variance Bivariate Distribution

Density Along $x$ and $y$-axis. shows only one mode
Density along the line $x = y$ shows the two modes of the bivariate distribution.
Modality: Equal Variance Bivariate Distribution

- Multivariate normal bimodal $\iff$ the univariate distribution along some line is bimodal.

- Extract the line of maximum separation.

- Use the bimodality condition in the univariate case.

For mixture equal variance normal: line joining the two means.
Modality: Equal Variance Bivariate Distribution

- Multivariate normal bimodal \( \iff \) the univariate distribution along some line is bimodal.
- Extract the line of maximum separation.
- Use the bimodality condition in the univariate case.

For mixture equal variance normal : line joining the two means.

**Theorem 1** *Multivariate Modality Condition: Equal variance Case*

\[
X \sim \pi N(\mu_1, \Sigma) + (1 - \pi) N(\mu_2, \Sigma)
\]

The distribution of \( X \) is bimodal iff

\[
(\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2) > 4
\]
Motivating Examples
Univariate Mixtures
Multivariate Mixtures: Equal Variance
Multivariate Mixtures: Unequal Variance
Bivariate Mixture Unequal variance
The Ridgeline Manifold
Higher Dimensions: No Contour plot available
Ridgeline Elevation Plot
Applications
Pi-plots and Analytical Solution
Modal Cluster
Hierarchical Modal Clustering with Example
Summary

Bivariate Mixture Unequal variance

Two components showing 3 modes
Motivating Examples

Univariate Mixtures

Multivariate Mixtures: Equal Variance

Multivariate Mixtures: Unequal Variance

Bivariate Mixture Unequal variance

The Ridgeline Manifold

Higher Dimensions: No Contour plot available

Ridgeline Elevation Plot

Applications

Pi-plots and Analytical Solution

Modal Cluster

Hierarchical Modal Clustering with Example

Summary

Line connecting the means
Exploring this line uncovers only two modes
Instead we should focus on the **ridge-line curve**
Density along the **ridge-line curve** uncovers the 3 modes.
The Ridgeline Manifold

\[ g(x) = \pi_1 \phi(x; \mu_1, \Sigma_1) + \pi_2 \phi(x; \mu_2, \Sigma_2) + \ldots + \pi_K \phi(x; \mu_K, \Sigma_K) \]

Mapping \( x^* : \mathbb{R}^{K-1} \rightarrow \mathbb{R}^D \)

\[ x^*(\alpha) = \left[ \alpha_1 \Sigma_1^{-1} + \alpha_2 \Sigma_2^{-1} + \ldots + \alpha_K \Sigma_K^{-1} \right]^{-1} \times \left[ \alpha_1 \Sigma_1^{-1} \mu_1 + \alpha_2 \Sigma_2^{-1} \mu_2 + \ldots + \alpha_K \Sigma_K^{-1} \mu_K \right] \]

- Image \( M \) [ridgeline surface or manifold.]
- \( K = 2 \), \( \implies \) one-dimensional ridgeline curve.

**Theorem 2** All the critical values of the \( D \)-dimensional multivariate mixture, and hence modes, antimodes and saddle points, lie in the manifold \( M \) given by \( x^*(\alpha) \)

- Enormous dimension reduction for \( d \gg K \).
  - Explore only \( K - 1 \) dimensions.
Two components, 3 dimensions, unequal variance:

\[
\mu_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \Sigma_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.05 \end{pmatrix},
\]

\[
\mu_2 = \begin{pmatrix} 1/\sqrt{2} \\ 2 \\ 1/\sqrt{2} \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 0.05 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}
\]

- No contour plot possible
- Advantages of dimension reduction
  - Visualization.
  - Reduced search space for finding modes.
Motivating Examples

Univariate Mixtures

Multivariate Mixtures: Equal Variance

Multivariate Mixtures: Unequal Variance

Bivariate Mixture Unequal variance

The Ridgeline Manifold

Higher Dimensions: No Contour plot available

Ridgeline Elevation Plot

Applications

Pi-plots and Analytical Solution

Modal Cluster

Hierarchical Modal Clustering with Example

Summary
There are 4 modes separated by 3 local minima in this elevation plot.
Motivating Examples
Univariate Mixtures
Multivariate Mixtures: Equal Variance
Multivariate Mixtures: Unequal Variance
Bivariate Mixture Unequal variance
The Ridgeline Manifold
Higher Dimensions: No Contour plot available
Ridgeline Elevation Plot
Applications
Pi-plots and Analytical Solution
Modal Cluster
Hierarchical Modal Clustering with Example
Summary
Assume that the 50 samples from the three species give the population parameters.

**Question:** Do they display one, two or three modes

\[ D = 4, \quad \rightarrow \quad \text{direct contour plotting of the density is not available.} \]

But as \( K = 3 \) modes live in 2 dimensions.

Examine the contour plot of this ridgeline manifold.
Motivating Examples

Univariate Mixtures

Multivariate Mixtures: Equal Variance

Multivariate Mixtures: Unequal Variance

Applications

Data Example: Iris Data

Iris Data: Ridgeline Contour Surface plot

Example: Egyptian Skull Data

Egyptian Skull Data: Ridgeline Contour plot

Pi-plots and Analytical Solution

Modal Cluster

Hierarchical Modal Clustering with Example

Summary
Example: Egyptian Skull Data

Egyptian Skull Data:
- Four measurements Maximal Breadth, Basibregmatic Height, Basialveolar Length, and Nasal Height of male Egyptian skulls
- 5 time periods (4000 BC, 3300 BC, 1850 BC, 200 BC, 150 AD).
- 30 skulls from each time period.

Again $D = 4$. No Contour plot available.

Here we analyze the 3 earliest time periods.
Motivating Examples

Univariate Mixtures

Multivariate Mixtures: Equal Variance

Multivariate Mixtures: Unequal Variance

Applications

Data Example: Iris Data

Example: Egyptian Skull Data

Pi-plots and Analytical Solution

Modal Cluster

Hierarchical Modal Clustering with Example

Summary
**Modality and mixing proportion: \( \Pi \)-plots**

**Ridgeline elevation/contour plots**
- Full information (location and height) of the modes and saddle-point.
- No dependence on \( \pi \). (Mixing proportions)

**\( \Pi \)-plots and Analytical Solutions**
- Only location of critical points.
- But dependence on the mixing parameter \( \pi \)

**Ridgeline elevation/contour plots**
- Full information (location and height) of the modes and saddle-point.
- No dependence on \( \pi \). (Mixing proportions)

**\( \Pi \)-plots and Analytical Solutions**
- Only location of critical points.
- But dependence on the mixing parameter \( \pi \)
The ‘\(\Pi\)-equation’

Two component case: If \(x^*(\alpha)\) is a critical value of \(h(\alpha)\) if it satisfies

\[
h'(\alpha) = \pi \phi_1(x^*(\alpha))' + \pi \phi_2(x^*(\alpha))' = 0, \text{ where } h(\alpha) = g(x_\alpha)
\]

Solving gives

\[
\Pi(\alpha) = \frac{\phi_2'(\alpha)}{\phi_2'(\alpha) - \phi_1'(\alpha)}.
\]

\(\Rightarrow\) if \(\alpha\) is a critical value then it solves the “pi-equation”:

\[
\Pi(\alpha) = \pi.
\]

- 1 solution \(\Rightarrow\) 1 mode
- 3 solution \(\Rightarrow\) 2 mode
- 5 solution \(\Rightarrow\) 3 mode
Motivating Examples

Univariate Mixtures

Multivariate Mixtures: Equal Variance

Multivariate Mixtures: Unequal Variance

Applications

Pi-plots and Analytical Solution

Modality and mixing proportion: \( \Pi \)-plots

The ‘\( \Pi \)-equation’

Example: \( \Pi \)-plot for unimodal density

Example: \( \Pi \)-plot for example with 4 modes

Analytical Solution for the number of modes

Modal Cluster

Hierarchical Modal Clustering with Example

Summary

Example: \( \Pi \)-plot for unimodal density

For any value of \( \pi \) we have only one crossing \( \Rightarrow \) unimodal.
Example: $\Pi$-plot for example with 4 modes

1, 3, 5 or 7 crossings $\implies$ 1, 2, 3 or 4 modes respectively.
Analytical Solution for the number of modes

\[ \kappa(\alpha) = [p(\alpha)]^2 [1 - \alpha \bar{\alpha} p(\alpha)], \]

where

\[ p(\alpha) = (\mu_2 - \mu_1)' \Sigma_1^{-1} S_{\alpha}^{-1} \Sigma_2^{-1} S_{\alpha}^{-1} \Sigma_1^{-1} S_{\alpha}^{-1} (\mu_2 - \mu_1). \]

where \( S_{\alpha} = \left[ \alpha_1 \Sigma_1^{-1} + \alpha_2 \Sigma_2^{-1} \right]^{-1} \left[ \alpha_1 \Sigma_1^{-1} \mu_1 + \alpha_2 \Sigma_2^{-1} \mu_2 \right] \)
Analytical Solution for the number of modes

\[ \kappa(\alpha) = [p(\alpha)]^2 [1 - \alpha \bar{\alpha} p(\alpha)], \]

where

\[ p(\alpha) = (\mu_2 - \mu_1)' \Sigma_1^{-1} S_{\alpha}^{-1} \Sigma_2^{-1} S_{\alpha}^{-1} \Sigma_1^{-1} S_{\alpha}^{-1} (\mu_2 - \mu_1). \]

where \( S_{\alpha} = \left[ \alpha_1 \Sigma_1^{-1} + \alpha_2 \Sigma_2^{-1} \right]^{-1} \left[ \alpha_1 \Sigma_1^{-1} \mu_1 + \alpha_2 \Sigma_2^{-1} \mu_2 \right] \)

Solution to:
- Quadratic for Equal Variance.
- Cubic for Proportional variance.

**Open Question**
- Upper bound for the number of modes for unequal variance case.
Associated mode of a point in the sample space

Any point $x$ can move along its path of steepest ascent, w.r.t. the density.

The associated mode of $x$ is that point after which no more ascent is possible.

For multivariate data focus on the **ridgeline manifold**.
Result. One can find the modes of a mixture density

\[ g(x) = \sum_k \pi_k f_k(x) \]

using a **Minorize-maximize (MM)** algorithm we can find the modes by iteratively solving the equations

\[ \sum \Pr(J = j|X = x_r) \nabla \log(f_j(x)) = 0 \]

for \( x_r \).
Mode Hunting in high dimensions

Result. One can find the modes of a mixture density

\[ g(x) = \sum_k \pi_k f_k(x) \]

using a **Minorize-maximize (MM)** algorithm we can find the modes by iteratively solving the equations

\[ \sum \Pr(J = j | X = x_r) \nabla \log(f_j(x)) = 0 \]

for \( x_r \).

- MM [Hunter and Lange] is not an “algorithm” but a prescription for constructing EM like algorithms.
- Surrogate function by majorizing or minorizing the objective function.
- Likelihood or missing data framework: not necessary
- General Convexity of function and relevant inequalities: important
Mode Hunting in high dimensions

Result. One can find the modes of a mixture density

\[ g(x) = \sum_k \pi_k f_k(x) \]

using a **Minorize-maximize (MM)** algorithm we can find the modes by iteratively solving the equations

\[ \sum \Pr(J = j | X = x_r) \nabla \log(f_j(x)) = 0 \]

for \( x_r \).

Normal mixture, common covariance \( \Sigma \),

\[ \implies x_{r+1} = \sum \Pr(J = j | X = x_r) \mu_j. \]

Posterior probability.

\[ \Pr(J = j | X = x_r) = \frac{\pi_j f_j(x_r)}{\sum_k \pi_k f_k(x_r)} \]
Properties of Modal Optimizer

- This iteration will, monotonely increase the density i.e.
  \[ g(x_{r+1}) \geq g(x_r). \]

- Discrete approximation to the **path of steepest ascent** until no further ascent is possible.

- Computationally simple steps even in **high-dimensions**
Pruning Bivariate Mixture components

5 components

Variance = 0.2*I
Pruning Bivariate Mixture components

5 components

Variance = 0.7*I
Pruning Bivariate Mixture components

5 components

Variance = 1.8*I
Pruning Bivariate Mixture components

5 components

Variance = 2.5*I
Motivating Examples

Univariate Mixtures

Multivariate Mixtures: Equal Variance

Multivariate Mixtures: Unequal Variance

Applications

Pi-plots and Analytical Solution

Modal Cluster
Associated mode of a point in the sample space
Mode Hunting in high dimensions
Properties of Modal Optimizer
Pruning Bivariate Mixture components

Clustering TBP-TAND interaction microarray data
Analysis of TBP-TAND interaction microarray data

Hierarchical Modal Clustering with Example

Summary

Clustering TBP-TAND interaction microarray data
Clustering TBP-TAND interaction microarray data

- 19-dim $\rightarrow$ Contour based visualization not possible
- Mixture Model based analysis result.
  - 12 components for equal variance.
  - 10 components for unequal variance.

4 modal clusters are consistent at a moderate scale.
Analysis of TBP-TAND interaction microarray data

- Competitor hierarchical and k-means for K=4 did not produce biologically meaningful structure.
- Initial analysis (by Pugh et.al.) done by hand “pruning” k-means for K=9.
- The 4 modal clusters have strong biological interpretation.

Details of Biological Interpretation: Ph D thesis Kate Huisinga, Penn State.
Hierarchical Modal Clustering

- Based on Kernel Density approach.
- Points belonging to the common mode of the density estimator defines a cluster.
- Varying the smoothing parameter provides multiscale analysis

Research Questions
- Interesting range of $h$ (kernel smoothing) to be explored.
- Incremental steps for $h$.
- Are the modes “significant.”
- Are the modes connected: ridge-line elevation plot.
Choice of $h$: Spectral Degrees of Freedom

- $sDOF$ analogous to $DOF$ in chi-squared goodness of fit [Details in Ray 2003, Ray and Lindsay 2006]

- Hypothesis testing $H_0 : G = F_\tau$:
  Fit statistic

  $$d_K(f, g) = \int_w \left( f_h^*(w) - g_h^*(w) \right)^2 dw$$

  $$\int K_{(h/2)}(x, w) dF(x) \quad \int K_{(h/2)}(x, w) dG(x)$$

- Choose $h$ such that $5 < sDOF < n/5$
If \( K(x, y) \):

\[
\int \int_S |K(x, y)|^2 dM(x) dM(y) < \infty
\]

\( K(x, y) \) generates a *Hilbert-Schmidt operator* on \( L_2(M) \) through the operation \((Kg)(x) = \int K(x, y)g(y) \, dM(y)\)

Theorem: Yosida (1980)

\[
K(x, y) = \sum_{j=1}^{\infty} \lambda_j \phi_j(x) \phi_j(y),
\]

\( \lambda_j \)'s eigenvalues \( \phi_j(x) \)'s are corresponding normalized eigen functions of \( K \) under baseline measure \( M \).
Asymptotically

$$d_K(f, g) \overset{d}{\to} \sum_{1}^{\infty} \lambda_j \chi^2_{(1)}$$

- Can be further approximated by a single chi-squared with df="Spectral Degrees of Freedom" $\chi^2_{sDOF}$
- Empirical estimate of $sDOF$ given by

$$\hat{sDOF} = \frac{(trace \hat{F}(K_{(h)_{c}}))^2}{trace \hat{F}(K^2_{(h)_{c}})} = \frac{\left[ \frac{1}{n} \sum_{i=1}^{n} K_{(h)_{c}}(x_i, x_i) \right]^2}{\frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j<i} \left( K_{(h)_{c}}(x_i, x_j) \right)^2}$$

where the centered kernel based on the empirical distribution is given by

$$K_{(h)_{c}}(x_i, x_j) = K_h(x_i, x_j) - \frac{1}{n} \sum_i K_h(x_i, x_j) - \frac{1}{n} \sum_i K_h(x_i, x_j) + \frac{1}{n^2} \sum_i \sum_j K_h(x_i, x_j)$$
Motivating Examples

Univariate Mixtures

Multivariate Mixtures: Equal Variance

Multivariate Mixtures: Unequal Variance

Applications

Pi-plots and Analytical Solution

Modal Cluster

Hierarchical Modal Clustering with Example

Hierarchical Modal Clustering

Choice of $h$: Spectral Degrees of Freedom

Spectral Decomposition

Satterthwaite Approximation and SDOF

Simulation: 4 component Normal

Simulation: The UI example

Modal Ridges: Gap test

Analysis of intensity data: modal clustering

Summary
Motivating Examples

Univariate Mixtures

Multivariate Mixtures: Equal Variance

Multivariate Mixtures: Unequal Variance

Applications

Pi-plots and Analytical Solution

Modal Cluster

Hierarchical Modal Clustering with Example

Choice of $h$: Spectral Degrees of Freedom

Spectral Decomposition

Satterthwaite Approximation and $SDOF$ Simulation: 4 component Normal

Simulation: The UI example

Modal Ridges: Gap test

Analysis of intensity data: modal clustering

Summary
Motivating Examples

Univariate Mixtures

Multivariate Mixtures: Equal Variance

Multivariate Mixtures: Unequal Variance

Applications

Pi-plots and Analytical Solution

Modal Cluster

Hierarchical Modal Clustering with Example

Choice of $h$: Spectral Degrees of Freedom

Spectral Decomposition

Satterthwaite Approximation and $SDOF$

Simulation: 4 component Normal

Simulation: The U I example

Modal Ridges: Gap test

Analysis of intensity data: modal clustering

Summary
Motivating Examples

Univariate Mixtures

Multivariate Mixtures: Equal Variance

Multivariate Mixtures: Unequal Variance

Applications

Pi-plots and Analytical Solution

Modal Cluster

Hierarchical Modal Clustering with Example

Hierarchical Modal Clustering

Choice of $h$: Spectral Degrees of Freedom

Spectral Decomposition

Satterthwaite Approximation and $S_DOF$

Simulation: 4 component Normal

Simulation: The U I example

Modal Ridges: Gap test

Analysis of intensity data: modal clustering

Summary
Simulation: The U I example

Normal Mixture based EM
Motivating Examples

Univariate Mixtures

Multivariate Mixtures: Equal Variance

Multivariate Mixtures: Unequal Variance

Applications

Pi-plots and Analytical Solution

Modal Cluster

Hierarchical Modal Clustering with Example

Hierarchical Modal Clustering

Choice of $h$: Spectral Degrees of Freedom

Spectral Decomposition

Satterthwaite Approximation and $SDOF$

Simulation: 4 component Normal

Simulation: The U I example

Modal Ridges: Gap test

Analysis of intensity data: modal clustering

Summary
Simulation: The U I example

Modal Clustering restricted to interesting range of \( h \)
Motivating Examples

Univariate Mixtures

Multivariate Mixtures: Equal Variance

Multivariate Mixtures: Unequal Variance

Applications

Pi-plots and Analytical Solution

Modal Cluster

Hierarchical Modal Clustering with Example

Hierarchical Modal Clustering

Choice of $h$: Spectral Degrees of Freedom

Spectral Decomposition

Satterthwaite Approximation and $SDOF$

Simulation: 4 component Normal

Simulation: The U I example

Modal Ridges: Gap test

Analysis of intensity data: modal clustering

Summary
Modal Ridges: Gap test

- Purpose: To find out the separateness of each pairwise mode.
- Evaluate the height along the ridgeline connecting the two modes.
- Modal ridge determined by the height of the saddle between the two modes.
- Use paired t-test of the kernel density estimates and multiple testing.
Motivating Examples

Univariate Mixtures

Multivariate Mixtures: Equal Variance

Multivariate Mixtures: Unequal Variance

Applications

Pi-plots and Analytical Solution

Modal Cluster

Hierarchical Modal Clustering with Example

Hierarchical Modal Clustering

Choice of $h$: Spectral Degrees of Freedom

Spectral Decomposition

Satterthwaite Approximation and SDOF

Simulation: 4 component Normal

Simulation: The U I example

Modal Ridges: Gap test

Analysis of intensity data: modal clustering

Summary

McGill University: Feb/March 2006 - slide #36
- Intensity sampled from inside and outside of kidney of humans using the m-rep model
- 39 CT images (in-plane)
- Intensity in: Adjusted Hounsfield Units
Motivating Examples

Univariate Mixtures

Multivariate Mixtures: Equal Variance

Multivariate Mixtures: Unequal Variance

Applications

Pi-plots and Analytical Solution

Modal Cluster

Hierarchical Modal Clustering with Example

Hierarchical Modal Clustering

Choice of $h$: Spectral Degrees of Freedom

Spectral Decomposition

Satterthwaite Approximation and $SDOF$

Simulation: 4 component Normal

Simulation: The U I example

Modal Ridges: Gap test

Analysis of intensity data: modal clustering

Summary
Analysis of intensity data: modal clustering

Multivariate data can arise from

- **fMRI Data** (3 dim), Measurement from different channels
  - T1, T2 and PD
- Texture Data (>8 dim)
- EEG Data: As large as 256 channels
Modal Clusters

- Combines the merits of model based and hierarchical clustering.
- Efficient even for arbitrary model and can be used in the context of non-parametric kernel based density estimation.
- For high-dimensional data, the ridge-lines and modes provide an appropriate dimension reduction.
- Software will be available soon.
Future Directions

- Dimension Reduction and visualization.
- Specifying the smoothing level, equivalent to the tree ascending levels in hierarchical clustering.
- Allowing outliers:
  - Mixture of Robust Distributions.
  - Associate modal classes only to certain % of data and allowing.
- Direction of maximal variation (discrimination) based on modal ridges.
Future Directions

- Dimension Reduction and visualization.
- Specifying the smoothing level, equivalent to the tree ascending levels in hierarchical clustering.
- Allowing outliers:
  - Mixture of Robust Distributions.
  - Associate modal classes only to certain % of data and allowing.
- Direction of maximal variation (discrimination) based on modal ridges.

Overcome the CURSE of Dimensionality
Future Directions

- Dimension Reduction and visualization.
- Specifying the smoothing level, equivalent to the tree ascending levels in hierarchical clustering.
- Allowing outliers:
  - Mixture of Robust Distributions.
  - Associate modal classes only to certain % of data and allowing.
- Direction of maximal variation (discrimination) based on modal ridges.

Goal: Explore the BLESSINGS of High Dimension
Collaborators

- **Mixture Models, Quadratic Distance, Modal Clusters**
  - Bruce G. Lindsay
  - Marianthi Markatou
  - Shu-Chuan (Grace) Chen
  - Jia Li

- **HDLSS geometry, bi-clustering, analysis of microarray data**
  - Steve Marron

- **Proteomics: Classification of MHC Binders, Vaccine Design, HMM**
  - Thomas B Kepler
  - Andrew Nobel
  - Scott Schmidler

- **Medical Imaging, Non-linear Shape Statistics, Imaging using tissue Mixtures**
  - Keith E. Muller
  - Stephen M. Pizer
  - Joshua Stough (Ph D student)
  - Ja Yeon Jeong (Ph D student)

- **Bayesian Model Selection, Generalized BIC, Structural Equations Model**
  - Jim Berger
  - Kenneth Bollen
  - Jane Zavisca
  - M. J. (Susie) Bayarri
Motivating Examples

Univariate Mixtures

Multivariate Mixtures: Equal Variance

Multivariate Mixtures: Unequal Variance

Applications

Pi-plots and Analytical Solution

Modal Cluster

Hierarchical Modal Clustering with Example

Summary

Concluding Remarks

Future Directions

Collaborators

References

Other interests: working papers

---

**Topography of Mixtures and Modal Clustering**


- Ray, S., Lindsay, B.G, Li, J. Hierarchical Modal Clustering. [Working Paper]

**Model Selection in High Dimension**

- Ray, S., Lindsay, B.G. Model selection in High-Dimensions: A Quadratic-risk Based Approach. [Ready for submission]

- Lindsay, B.G., Ray, S., Chen, S.C., Yang, K., Markatou M. Quadratic distances on probabilities: the foundations.[Submitted]

- Lindsay, B.G., Ray, S., Chen, S.C., Yang, K., Markatou M. Quadratic distances on probabilities: Multivariate construction using tuneable diffusion kernels. [Tech report]

**HDLSS and Microarray Analysis**

- Ray, S., Marron, J.S. Feature selection in High dimensional Low sample size data using two-way mixtures.
Other interests: working papers

Proteomics: Immunoinformatics


Empirical Bayes: Model Selection

- Ray, S., Bollen, K.A. Bayesian model selection criteria for hierarchical and structural equation models. (Invited Book Chapter, Editor: Dunson, D.)