

High to High Dimensional Multivariate Mixture Regression

Alex White

4/27/2021

Motivation

Understanding how DNA methylation affects/predicts gene expression in heterogeneous patient populations.

Suppose for some patient population we have:

- ▶ DNA methylation levels for some set of CpG sites
- ▶ Gene expression levels for some clinically relevant sites.

We are interested in:

1. Clustering patients into the correct sub-group
2. Within sub-group, finding relationship between CpG site & gene expression

Idea

Goal: Correctly cluster observations & regress in high dimensional X & Y .

- ▶ $Y_{n \times q}$ Matrix of responses
- ▶ $X_{n \times p}$ Design matrix
- ▶ $A_{p \times q}$ Coefficient matrix (sparse in p)
- ▶ k clusters
- ▶ Assume (for now) response covariance (Σ_k) is diagonal
 - ▶ Strong yet common assumption due to lack of degrees of freedom for correct variance specification

$$f(\mathbf{y}_i \mid \mathbf{x}_i; \theta) = \sum_{k=1}^K \pi_k \mathcal{N}_q(\mathbf{y}_i; \mathbf{x}_i A_k, \Sigma_k)$$

- ▶ Parameter space $\theta = \{\pi_k, A_k, \Sigma_k; k = 1 \dots K\}$ solved by general EM using SARRS to compute A_k .

Multivariate Regression

- Used when there are multiple response variables

$$\begin{pmatrix} y_{11} & y_{12} & \dots & y_{1p} \\ y_{21} & y_{22} & \dots & y_{2p} \\ y_{31} & y_{32} & \dots & y_{3p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \dots & y_{np} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1q} \\ 1 & x_{21} & x_{22} & \dots & x_{2q} \\ 1 & x_{31} & x_{32} & \dots & x_{3q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nq} \end{pmatrix} \begin{pmatrix} \beta_{01} & \beta_{02} & \dots & \beta_{0p} \\ \beta_{11} & \beta_{12} & \dots & \beta_{1p} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{q1} & \beta_{q2} & \dots & \beta_{qp} \end{pmatrix} + \begin{pmatrix} \epsilon_{11} & \epsilon_{12} & \dots & \epsilon_{1p} \\ \epsilon_{21} & \epsilon_{22} & \dots & \epsilon_{2p} \\ \epsilon_{31} & \epsilon_{32} & \dots & \epsilon_{3p} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_{n1} & \epsilon_{n2} & \dots & \epsilon_{np} \end{pmatrix}$$

Figure 1: “Multivariate Regression”

Mixed Regression

- Used when there are multiple response variables

$$\begin{pmatrix} y_{11} & y_{12} & \dots & y_{1p} \\ y_{21} & y_{22} & \dots & y_{2p} \\ y_{31} & y_{32} & \dots & y_{3p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \dots & y_{np} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1q} \\ 1 & x_{21} & x_{22} & \dots & x_{2q} \\ 1 & x_{31} & x_{32} & \dots & x_{3q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nq} \end{pmatrix} \begin{pmatrix} \beta_{01} & \beta_{02} & \dots & \beta_{0p} \\ \beta_{11} & \beta_{12} & \dots & \beta_{1p} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{q1} & \beta_{q2} & \dots & \beta_{qp} \end{pmatrix} + \begin{pmatrix} \epsilon_{11} & \epsilon_{12} & \dots & \epsilon_{1p} \\ \epsilon_{21} & \epsilon_{22} & \dots & \epsilon_{2p} \\ \epsilon_{31} & \epsilon_{32} & \dots & \epsilon_{3p} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_{n1} & \epsilon_{n2} & \dots & \epsilon_{np} \end{pmatrix}$$

Figure 2: “Mixed Regression”

Reduced Rank Regression

- ▶ Coefficient matrix sparse in q dimension
- ▶ Only few genes are associated with methylation

$$\begin{pmatrix} y_{11} & y_{12} & \dots & y_{1p} \\ y_{21} & y_{22} & \dots & y_{2p} \\ y_{31} & y_{32} & \dots & y_{3p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \dots & y_{np} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1q} \\ 1 & x_{21} & x_{22} & \dots & x_{2q} \\ 1 & x_{31} & x_{32} & \dots & x_{3q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nq} \end{pmatrix} \begin{pmatrix} \beta_{01} & \beta_{02} & \dots & \beta_{0p} \\ \beta_{11} & \beta_{12} & \dots & \beta_{1p} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{q1} & \beta_{q2} & \dots & \beta_{qp} \end{pmatrix} + \begin{pmatrix} \epsilon_{11} & \epsilon_{12} & \dots & \epsilon_{1p} \\ \epsilon_{21} & \epsilon_{22} & \dots & \epsilon_{2p} \\ \epsilon_{31} & \epsilon_{32} & \dots & \epsilon_{3p} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_{n1} & \epsilon_{n2} & \dots & \epsilon_{np} \end{pmatrix}$$

Figure 3: “Reduced Rank Regression”

HTH Mixture Algorithm

- ▶ Initialize: $\pi_k^{(0)} = \frac{n_k^{(0)}}{n}$
- ▶ Randomly initialize observations into k clusters

While not converged ($m = 1, \dots, M$) do:

- ▶ for $k = 1, \dots, K$ apply SARRS on all observations in $C_k^{(m-1)}$ to obtain $A_k^{(m)}, \Sigma_k^{(m)}$
- ▶ compute $\mu_{ik}^{(m)} = \mathcal{N}_p(\mathbf{y}_i; A_k^{(m)} \mathbf{x}_i, \Sigma_k^{(m)})$
- ▶ $C_k^{(m)} = \{i | \text{ML component } k\}$
- ▶ $\pi_k^{(m)} = \frac{n_k^{(m)}}{n}$

HTH Mixture Algorithm

- ▶ Empirically, HTH Mixture reaches local maximum quickly
- ▶ Need to determine method for efficiently exploring the likelihood space with different random initialization.

Likelihood Space Exploration Idea 1

- ▶ Mimic the idea of chains from MCMC
- ▶ In parallel, run large number of chains and select the one which reaches the largest likelihood.
- ▶ This method works well, although not the most efficient way to explore the space.

Likelihood Space Exploration Idea 2

- ▶ To more efficiently explore the space, when a local maximum is reached, perturb the initialized state by a proportion p (e.g. 20%) and rerun.

Data Simulation

- ▶ X_k consists of iid random vectors sample from $MVN(\mathbf{0}, \Sigma_k)$
- ▶ Σ_k diagonal
- ▶ Noise matrix $Z_k \in \mathbb{R}^{n \times q}$ has iid $N(0, \sigma^2)$ entries
- ▶ $A_k = \begin{pmatrix} b_k B_{0_k} B_{1_k} \\ 0 \end{pmatrix}$
 - ▶ with $b > 0$, $B_0 \in \mathbb{R}^{s \times r}$, $B_1 \in \mathbb{R}^{r \times q}$
- ▶ $Y_k = X_k A_k + Z_k$

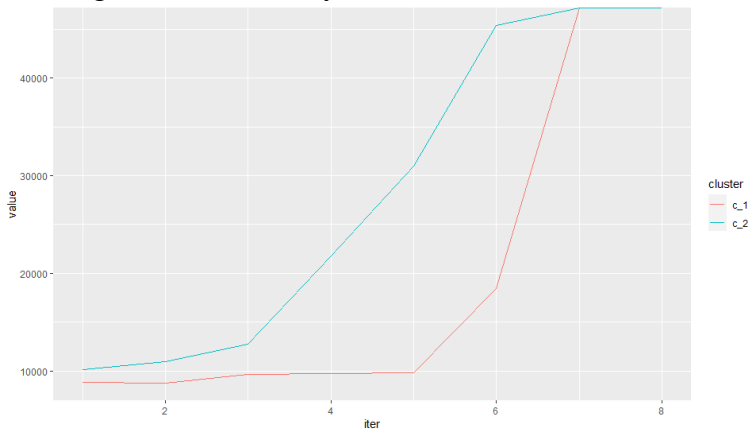
Finally, combine X & Y

Simulation

- ▶ N: 50 – 400
- ▶ s: 5 – 20
- ▶ rank: 1 – 2
- ▶ B: 1
- ▶ $p = m = 1000$
- ▶ Perturb: 20%
- ▶ Cluster with equal probability
- ▶ K: 2
- ▶ σ : 1
- ▶ CHains: 500

Results

- HTH algorithm consistently reaches local maximum



Results

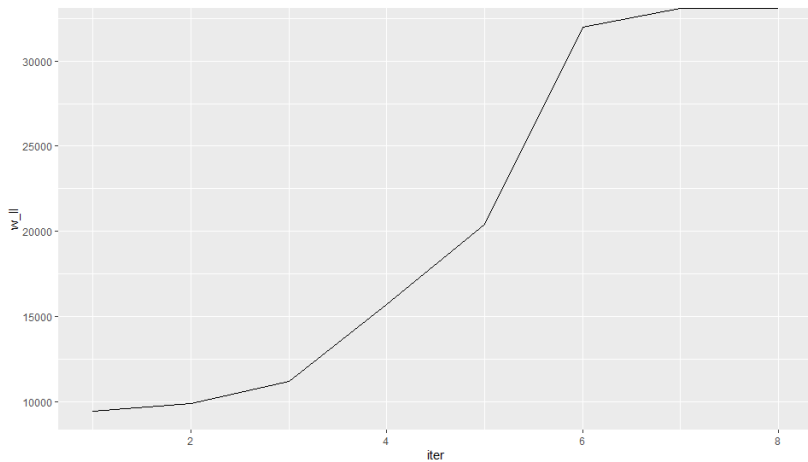
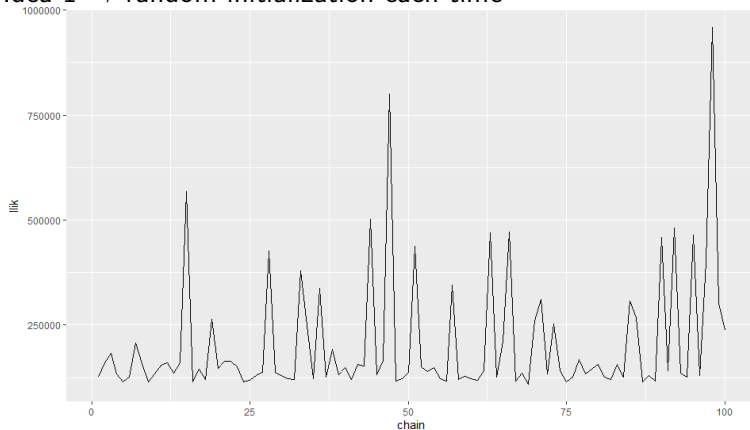


Figure 4: Weighted Likelihood

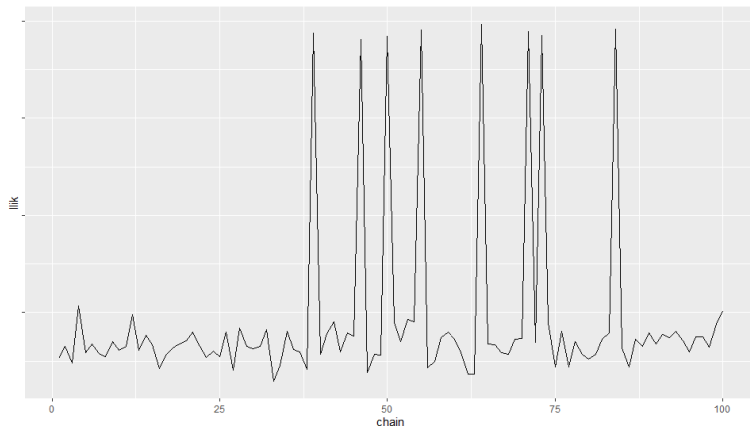
Results

- Idea 1 → random initialization each time



Results

- ▶ Idea 2 \rightarrow perturb local maximum
- ▶ Appears to be more efficient



Results

N	s	R	hthmix	kmeans
50	5	1	0.90	0.70
100	5	1	1.00	0.66
200	5	1	1.00	0.51
400	5	1	1.00	0.54
50	10	1	0.84	0.62
100	10	1	0.92	0.50
200	10	1	1.00	0.66
400	10	1	1.00	0.52
50	20	1	0.80	0.72
100	20	1	0.80	0.53
200	20	1	0.98	0.68
400	20	1	1.00	0.51

Results

N	s	R	hthmix	kmeans
50	5	2	1.00	0.64
100	5	2	1.00	0.72
200	5	2	1.00	0.52
400	5	2	1.00	0.64
50	10	2	0.86	0.58
100	10	2	1.00	0.55
200	10	2	1.00	0.53
400	10	2	1.00	0.68
50	20	2	0.84	0.64
100	20	2	0.78	0.53
200	20	2	1.00	0.71
400	20	2	1.00	0.65

Performance

- ▶ In simulated data, current algorithm clusters well (perfectly in most cases given enough chains)
- ▶ Challenges:
 - ▶ Likelihood space is explored inefficiently, sometimes requires large number of chains
 - ▶ Computationally cumbersome/inefficient