Alternating direction method of multipliers (ADMM) algorithm for solving bi-clustering problem: application to microbiome data analysis

Lanqiu Yao

NYU School of Medicine
Division of Biostatistics

ly1192@nyu.edu

May 8, 2019

Overview

Introduction: Clustering

Clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters)

Pattern recognition

Cells: Prokaryotic cells, Eukaryotic cells...

Lung cancer: Lung adenocarcinoma, Squamous cell carcinoma, Small cell carcinoma,. Large cell carcinoma ...

Restaurant: good restaurant, bad restaurant.

Method:

Hierarchical clustering

K-means

...

Introduction: Biclustering

The previously mentioned clustering methods can only identify one kinds of features.

Some times people want to cluster the observations and features simultaneously. Then biclustering algorithms are needed.

smoking/nonsomking

lung cancer/health

Biclustering has applications in a wide world. It can help people identify subsets of genes/ microbiome within subsets of conditions.
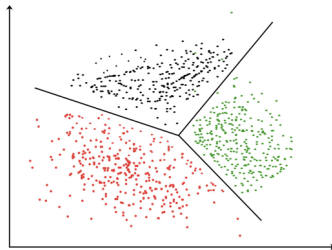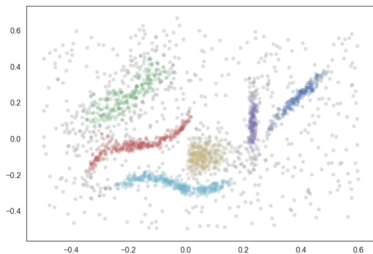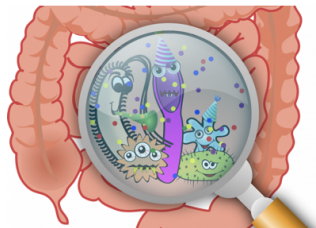
The Effect of Diet on the Human Gut Microbiome [1]

Low-fat, plant polysaccharide–rich diet

High-sugar "Western" diet

Human Gut Microbiome

---

[1] Turnbaugh, Peter J., et al. "The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice." Science translational medicine 1.6 (2009): 6ra14-6ra14.

ADMM

ADMM

Alternating direction method of multipliers(ADMM) is an algorithm that is intended to blend the decomposability of dual ascent with the superior convergence properties of the method of multipliers. The algorithm solves problems in the form

$$\min f(x) + g(z)$$

$$Ax + Bz = 0$$

with variable $x \in R^n$, $z \in R^m$, where $A \in R^{pn}$, $B \in R^{p \times m}$, and $C \in R^p$

Let's look at the biclustering ADMM algorithm

ADMM

X is an $n \times p$ matrix, with n observations and p features.

$$\mathbf{X}_{n,p} = \left[\begin{array}{cccc} X_{11} & X_{12} & ... & X_{1p} \\ X_{21} & X_{22} & ... & X_{2p} \\ ... & ... & ... & ... \\ X_{n1} & X_{n2} & ... & X_{np} \end{array} \right] = (X_{\cdot 1}, X_{\cdot 2}, ..., X_{\cdot p})$$

Assume that the n observations belong to R unknown and non-overlape classes.

Assume that the p features belong to C unknown and non-overlap classes.

Cluster mean: $\mu = \{\mu_{r,c}\} \in \mathbb{R}^{R \times C}$

Cluster mean: $\mu_{r,c} = \frac{1}{|R||C|} \sum_{i \in R, j \in C} x_{ij}$

Let $A = \{a_{ij}\} \in \mathbb{R}^{n \times p}$, where A matrix is the estimated the cluster labels/pattern of matrix X

ADMM

The convex biclustering problem can be formulated as:

$$\min_{A \in \mathbb{R}^{n \times p}} \frac{1}{2} \sum_{i=1}^{p} ||X_{\cdot i} - A_{\cdot i}||_2^2 + \gamma_1 \sum_{l \in E_1} w_l ||\mathbf{v}_l||_q + \gamma_2 \sum_{k \in E_2} u_k ||\mathbf{z}_k||_q$$

$$s.t. \ A_{l_1 \cdot} - A_{l_2 \cdot} - \mathbf{v}_l = 0$$
$$A_{\cdot k_1} - A_{\cdot k_2} - \mathbf{z}_k = 0 \tag{*}$$

where,

$E_1 = \{l = (l_1, l_2) : 1 \le l_1 < l_2 \le n\}$

$E_2 = \{k = (k_1, k_2) : 1 \le k_1 < k_2 \le n\}$

$|| \cdot ||_q$ is the L_q norm, q can be $1, 2, ...,$ or ∞

ADMM

Let's see an example:

$$\mu_{1,1} = 1.5, \mu_{1,2} = -2, \mu_{2,1} = -1.5, \mu_{2,2} = 2$$

ADMM

Let's see an example:
$\mu_{1,1} = 1.5, \mu_{1,2} = -2, \mu_{2,1} = -1.5, \mu_{2,2} = 2$

X			
1.4	1.6	-2	-2.5
1.7	1.3	-1.5	-2
-1.4	-1.6	2.1	1.9
-1.7	-1.3	1.8	2.2

1.4	1.6	-2	-2.5
1.7	1.3	-1.5	-2
-1.4	-1.6	2.1	1.9
-1.7	-1.3	1.8	2.2

A			
1.5	1.5	-2	-2
1.5	1.5	-2	-2
-1.5	-1.5	2	2
-1.5	-1.5	2	2

ADMM

To solve this constrained optimization problem we can use augmented Lagrangian method:

$$
\begin{aligned}
\mathcal{L}_{\nu_1,\nu_2}(A, V, Z, \Lambda_1, \Lambda_2) = & \frac{1}{2}\sum_{i=1}^{p} ||X_{\cdot i} - A_{\cdot i}||_2^2 + \gamma_1 \sum_{l \in E_1} w_l ||\mathbf{v}_l||_q \\
& + \gamma_2 \sum_{k \in E_2} u_k ||\mathbf{z}_k||_q + \\
& \sum_{l \in E_1} < \lambda_{1l}, \mathbf{v}_l - A_{l_1 \cdot} + A_{l_2 \cdot} > \\
& + \frac{\nu_1}{2} \sum_{l \in E_1} ||\mathbf{v}_l - A_{l_1 \cdot} + A_{l_2 \cdot}||_2^2 \\
& + \sum_{k \in E_2} < \lambda_{2k}, \mathbf{z}_k - A_{\cdot k_1} + A_{\cdot k_2} > \\
& + \frac{\nu_2}{2} \sum_{k \in E_2} ||\mathbf{z}_k - A_{\cdot k_1} + A_{\cdot k_2}||_2^2
\end{aligned}
$$

ADMM

Parameters:

$\nu_1, \nu_1, \gamma_1, \gamma_2$

Need to solve: A

Iterations:

Update A

Update $\mathbf{V}, \mathbf{Z}$

Update Λ_1, Λ_2

ADMM - Update A

It is the most challenging part. We need to minimize

$$
\begin{aligned}
f(A) =& \frac{1}{2} \sum_{i=1}^{p} ||X_{\cdot i} - A_{\cdot i}||_2^2 \\
&+ \frac{\nu_1}{2} \sum_{l \in E_1} ||\tilde{\mathbf{v}}_l - A_{l_1 \cdot} + A_{l_2 \cdot}||_2^2 \\
&+ \frac{\nu_2}{2} \sum_{k \in E_2} ||\tilde{\mathbf{z_k}} - A_{\cdot k_1} + A_{\cdot k_2}||_2^2
\end{aligned}
$$

where,

$\tilde{\mathbf{v}}_l = \mathbf{v}_l + \frac{1}{\nu_1} \lambda_{1l}$

$\tilde{\mathbf{z_k}} = \mathbf{z_k} + \frac{1}{\nu_2} \lambda_{2k}$

ADMM - Update A

We could calculate its derivation. And the system of equations is finally equivalent to a Sylvester equation:

$$MA + AN = C \tag{1}$$

where,

$$M = I_n + \nu_1 \sum_{l \in E_1} (e_{l_1} - e_{l_2})(e_{l_1} - e_{l_2})^T \tag{2}$$

$$N = \nu_2 \sum_{k \in E_2} (e_{k_1}^* - e_{k_2}^*)(e_{k_1}^* - e_{k_2}^*)^T \tag{3}$$

$$C = X + \sum_{l in_1} (e_{l_1} - e_{l_2})(\lambda_{1l} + \nu_1 \mathbf{v_l})^T + (\lambda_{2k} + \nu_2 \mathbf{z_k})(e_{k_1}^* - e_{k_2}^*)^T \tag{4}$$

where $A_{l_1 \cdot} + A_{l_2 \cdot} = A(e_{l_1} - e_{l_2})$, e_{l_1} and e_{l_2} are $p \times 1$ vectors.
And similarly, $A_{\cdot k_1} + A_{\cdot k_2} = A^T(e_{k_1}^* - e_{k_2}^*)$, $e_{k_1}^*$ and $e_{k_2}^*$ are $n \times 1$ vectors.

Algorithm for solving Sylvester equation: BartelsStewart algorithm [2]

Input data: $\mathbf{A}_1 \in \mathbb{R}^{m \times m}$, $\mathbf{A}_2 \in \mathbb{R}^{n \times n}$, $\mathbf{D} \in \mathbb{R}^{m \times n}$
Output data: solution $\mathbf{X} \in \mathbb{R}^{m \times n}$

(1) Compute real Schur decomposition of $\mathbf{A}_1$, $\mathbf{A}_1 = \mathbf{Q}_1 \mathbf{R}_1 \mathbf{Q}_1^T$.

(2) If $\mathbf{A}_2 = \mathbf{A}_1$, set $\mathbf{Q}_2 = \mathbf{Q}_1$, $\mathbf{R}_2 = \mathbf{R}_1$;

    else if $\mathbf{A}_2 = \mathbf{A}_1^T$, get $\mathbf{Q}_2$, $\mathbf{R}_2$ from $\mathbf{Q}_1$, $\mathbf{R}_1$ as follows:
    $idx = [\text{size}(\mathbf{A}_1, 1) : -1 : 1]$; $\mathbf{Q}_2 = \mathbf{Q}_1(:, idx)$;
    $\mathbf{R}_2 = \mathbf{R}_1(idx, idx)^T$;

    else, compute real Schur decomposition of $\mathbf{A}_2$, $\mathbf{A}_2 = \mathbf{Q}_2 \mathbf{R}_2 \mathbf{Q}_2^T$.

(3) $\mathbf{D} \leftarrow \mathbf{Q}_1^T \mathbf{D} \mathbf{Q}_2$, $Rsq \leftarrow \mathbf{R}_1 * \mathbf{R}_1$, $\mathbf{I} \leftarrow eye(m)$, $j \leftarrow 1$.

(4) While $(j < n + 1)$

    if $j < n$ and $\mathbf{R}_2(j+1, j) < 10 * \epsilon * \max(|\mathbf{R}_2(j,j)|, |\mathbf{R}_2(j+1, j+1)|)$
    (a) $\mathbf{b} \leftarrow -\mathbf{D}(:, j) - \mathbf{X}(:, 1 : j-1) * \mathbf{R}_2(1 : j-1, j)$;
    (b) solve the linear equations $(\mathbf{R}_1 + \mathbf{R}_2(j,j)\mathbf{I})\mathbf{x} = \mathbf{b}$ for $\mathbf{x}$, set $\mathbf{X}(:, j) \leftarrow \mathbf{x}$;
    (c) $j \leftarrow j+1$
    else
    (a) $r_{11} \leftarrow \mathbf{R}_2(j,j)$, $r_{12} \leftarrow \mathbf{R}_2(j, j+1)$,
    $r_{21} \leftarrow \mathbf{R}_2(j+1, j)$, $r_{22} \leftarrow \mathbf{R}_2(j+1, j+1)$;
    (b) $\mathbf{b} \leftarrow -\mathbf{D}(:, j : j+1) - \mathbf{X}(:, 1 : j-1) * \mathbf{R}_2(1 : j-1, j : j+1)$;
    $\mathbf{b} \leftarrow [\mathbf{R}_1 \mathbf{b}(:, 1) + r_{22}\mathbf{b}(:, 1) - r_{21}\mathbf{b}(:, 2), \mathbf{R}_1 \mathbf{b}(:, 2) + r_{11}\mathbf{b}(:, 2) - r_{12}\mathbf{b}(:, 1)]$
    (c) block solve the linear equations
    $(Rsq + (r_{11} + r_{22})\mathbf{R}_1 + (r_{11}r_{22} - r_{12}r_{21})\mathbf{I})\mathbf{x} = \mathbf{b}$ for $\mathbf{x}$, set $\mathbf{X}(:, j : j+1) \leftarrow \mathbf{x}$;
    (d) $j \leftarrow j+2$
    end if.

(5) The solution $\mathbf{X}$ in the original basis is: $\mathbf{X} \leftarrow \mathbf{Q}_1 \mathbf{X} \mathbf{Q}_2^T$.

---

[2] Sorensen, Danny C., and Yunkai Zhou. "Direct methods for matrix Sylvester and Lyapunov equations." Journal of Applied Mathematics 6.2003 (2003): 277-303.

Proximal maps [3]

$$\mathbf{v}_l = \text{argmin}_{\mathbf{v}_l} \frac{1}{2}||\mathbf{v}_l - (A_{l_1\cdot} - A_{l_2\cdot} + \nu_1^{-1}\lambda_{1l})||_2^2 + \frac{\gamma_1 w_l}{\nu_1}||\mathbf{v}_l||_q$$

$$= prox_{\delta_{1l}||\cdot||_q}(A_{l_1\cdot} - A_{l_2\cdot} + \nu_1^{-1}\lambda_{1l})$$

$$\mathbf{z}_k = \text{argmin}_{\mathbf{z}_k} \frac{1}{2}||\mathbf{z}_k - (A_{\cdot k_1} - A_{\cdot k_2} + \nu_2^{-1}\lambda_{2k})||_2^2 + \frac{\gamma_2 u_k}{\nu_2}||\mathbf{z}_k||_q$$

$$= prox_{\delta_{2k}||\cdot||_q}(A_{\cdot k_1} - A_{\cdot k_2} + \nu_2^{-1}\lambda_{2k})6$$

Norm	$\Omega(\mathbf{v})$	$\text{prox}_{\sigma\Omega}(\mathbf{v})$	Comment
ℓ_1	$\|\mathbf{v}\|_1$	$\left[1 - \frac{\sigma}{\|v_l\|}\right]_+ v_l$	Elementwise soft-thresholding
ℓ_2	$\|\mathbf{v}\|_2$	$\left[1 - \frac{\sigma}{\|\mathbf{v}\|_2}\right]_+ \mathbf{v}$	Blockwise soft-thresholding
ℓ_∞	$\|\mathbf{v}\|_\infty$	$\mathbf{v} - \mathcal{P}_{\sigma S}(\mathbf{v})$	S is the unit simplex

[3]Chi, Eric C., and Kenneth Lange. "Splitting methods for convex clustering." Journal of Computational and Graphical Statistics 24.4 (2015): 994-1013.

It is easy to update λ_{1l} and λ_{2k}

$$\lambda_{1l} = \lambda_{1l} + \nu_1(\mathbf{v}_l - A_{l_1\cdot} + A_{l_2\cdot}) \tag{7}$$

$$\lambda_{2k} = \lambda_{2k} + \nu_2(\mathbf{z}_k - A_{\cdot k_1} + A_{\cdot k_2}) \tag{8}$$

ADMM

Algorithm ADMM

---

0. Set parameters: $\gamma_1, \gamma_2, \nu_1, \nu_2$
1. Initialize A_0, $\mathbf{v_0}$, $\mathbf{z_0}$
2. Update A
$$M = I_n + \nu_1 \sum_{l \in E_1}(e_{l_1} - e_{l_2})(e_{l_1} - e_{l_2})^T$$
$$N = \nu_2 \sum_{k \in E_2}(e^*_{k_1} - e^*_{k_2})(e^*_{k_1} - e^*_{k_2})^T$$
$$C = X + \sum_{l \in E_1}(e_{l_1} - e_{l_2})(\lambda_{1l} + \nu_1 \mathbf{v_l})^T + (\lambda_{2k} + \nu_2 \mathbf{z_k})(e^*_{k_1} - e^*_{k_2})^T$$
Solve: $MA + AN = C$
3. Update $\mathbf{v}$, $\mathbf{z}$
$$\mathbf{v}_l = prox_{\delta_{1l}||\cdot||_q}(A_{l_1 \cdot} - A_{l_2 \cdot} + \nu_1^{-1}\lambda_{1l})$$
$$\mathbf{z}_k = prox_{\delta_{2k}||\cdot||_q}(A_{\cdot k_1} - A_{\cdot k_2} + \nu_2^{-1}\lambda_{2k})$$
4. Update λ_1, λ_2
$$\lambda_{1l} = \lambda_{1l} + \nu_1(\mathbf{v}_l - A_{l_1 \cdot} + A_{l_2 \cdot})$$
$$\lambda_{2k} = \lambda_{2k} + \nu_2(\mathbf{z}_k - A_{\cdot k_1} + A_{\cdot k_2})$$
5. Repeat 2-4 until coverage.

Compare ADMM to other bi-clustering method.
Here, I would like to compare it with Convex BiclusteRing Algorithm
(COBRA)[4]

[4]Chi, Eric C., and Kenneth Lange. "Splitting methods for convex clustering." Journal of Computational and Graphical Statistics 24.4 (2015): 994–1013.

COBRA

Convex BiclusteRing Algorithm (COBRA)

The method also focused on solving the Lagrangian equation.
But instead of solving the Sylvester equation, an approximate method is applied.

$$MA + AN = C \tag{1}$$

That is, when updating A, ν_1 and ν_2 are temporally set as 0. Then

$$M = I_n, N = 0$$

$$A = C = X + \sum_{lin_1}(e_{l_1} - e_{l_2})(\lambda_{1l} + \nu_1 \mathbf{v_l})^T + (\lambda_{2k} + \nu_2 \mathbf{z_k})(e_{k_1}^* - e_{k_2}^*)^T$$

Complexity

The complexity of ADMM.

For each iteration:

1. Update A

    Calculate M, N, C: $\mathcal{O}(n)$

    Solve Sylvester equation by BartelsStewart algorithm[5]: $\mathcal{O}(n^3)$

2. Update $\mathbf{v}, \mathbf{z}$

    $\mathcal{O}(n)$

3. Update $\Lambda_{1,2}$

    $\mathcal{O}(n)$

Therefore the time complexity is $\mathcal{O}(n^3)$ for each iteration.

The iteration numbers are usually not sure. But the algorithm will converage finally in the end.

---

[5]Sorensen, Danny C., and Yunkai Zhou. "Direct methods for matrix Sylvester and Lyapunov equations." Journal of Applied Mathematics 6.2003 (2003): 277-303.

Results

Results:

Simulation

Microbiome data

Simulation

Data generation:

$n \times p = 80 \times 60$

4 clusters: 2 row clusters and 2

cluster means: $\mu = [1, 2, -1, -2]$
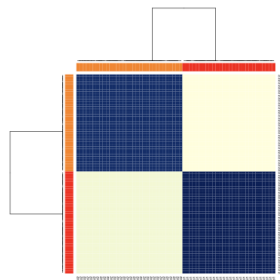
cells: $x_i j \sim N(\mu, \sigma)$, $\sigma = 1$
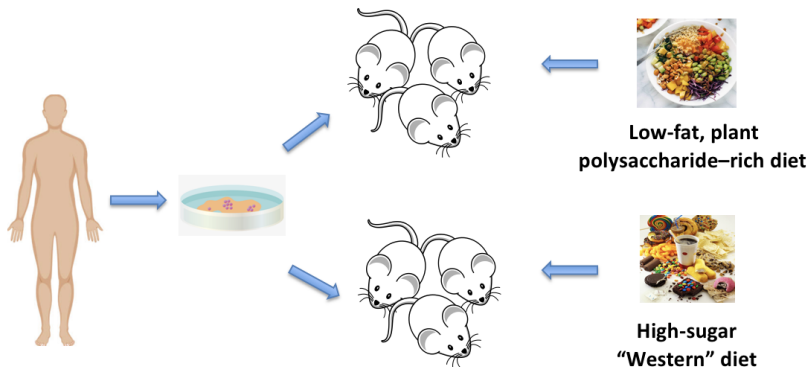
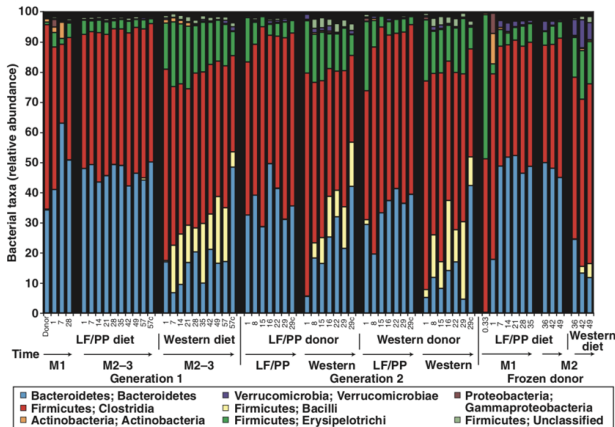Results - Heatmaps
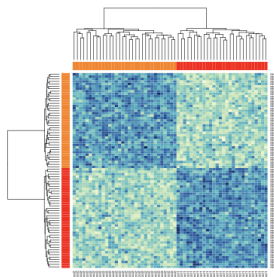
Raw Data
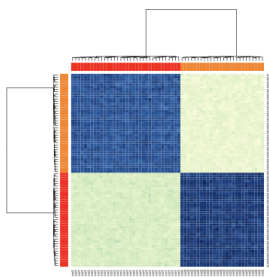
COBRA

ADMM

The brief experiment:

Low-fat, plant polysaccharide–rich diet

High-sugar "Western" diet

The effect of diet on the microbiome?

Microbiome

After 4 weeks, the sequencing of the 16S ribosomal RNA (rRNA) genes was done.

The bacterial taxa relative abundance is showing below:

Microbiome

To make it easy, I chose the two most abundance bacterial family:
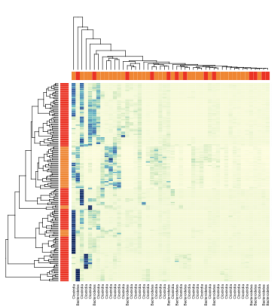
Clostridia

Bacteroidetes

In their dataset, $X = 172 \times 48$
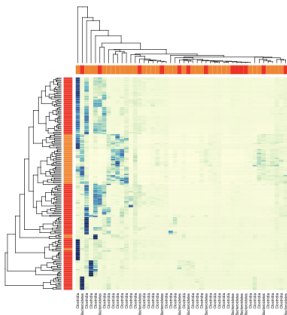172 mice and 48 kinds of bacteria that belong to the above two families.

Family	Freq
Bacteroidetes	13
Clostridia	35

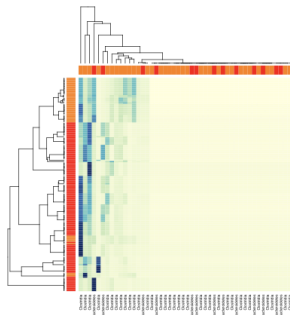| Diet	Freq
Low-fat, plant polysaccharide–rich diet	127
High-sugar "Western" diet	45

Microbiome

Results - Heatmaps

Raw Data   COBRA   ADMM

Conclusion

Biclustering algorithms can cluster the observations and features simultaneously.

ADMM algorithm works better than COBRA since it get the exact solution of the augmented Lagrangian problem while COBRA gives an approximate estimator.

The selection of the tuning parameters in convex clustering is important

Time complexity

Reference

[1] Turnbaugh, Peter J., et al. "The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice." Science translational medicine 1.6 (2009): 6ra14-6ra14.

[2] Chi, Eric C., Genevera I. Allen, and Richard G. Baraniuk. "Convex biclustering." Biometrics 73.1 (2017): 10-19.

[3] Sorensen, Danny C., and Yunkai Zhou. "Direct methods for matrix Sylvester and Lyapunov equations." Journal of Applied Mathematics 6.2003 (2003): 277-303.

[4] Chi, Eric C., and Kenneth Lange. "Splitting methods for convex clustering." Journal of Computational and Graphical Statistics 24.4 (2015): 994-1013.

[5] Chi, Eric C., and Kenneth Lange. "Splitting methods for convex clustering." Journal of Computational and Graphical Statistics 24.4 (2015): 994-1013.

Thank you!

ADMM