Rutgers Business School
Newark and New Brunswick

SDP Approximation of K-means Clustering

Reference: Approximating K-means-type clustering via semidefinite programming, Jiming Peng & Yu Wei

Yuan Qu

Main Idea

RUTGERS
Rutgers Business School
Newark and New Brunswick

- Clustering: assign n points into k clusters, based on the minimal sum-of-squares (MSSC)
- MSSC —> 0-1 SDP
- Extract feasible solution of 0-1 SDP

Clustering

Rutgers Business School
Newark and New Brunswick

- Given a set S of n points in d-dimensional Euclidean space.

$$S = \{s_i = (s_{i1}, s_{i2}, \ldots, s_{id})^T \in R^d,\ i = 1,2,\ldots,n\}$$

- k disjoint clusters $M = (M_1, M_2, \ldots, M_k)$ centered at cluster center (centroid) c_j based on the total sum-of-squares

$$f(S, M) = \sum_{j=1}^{k} \sum_{i=1}^{|M_j|} \|s_i^{(j)} - c_j\|^2$$

- $s_i^{(j)}$ — the i-th point in M_j

Clustering

$$f(S, M) = \sum_{j=1}^{k} \sum_{i=1}^{|M_j|} \| s_i^{(j)} - c_j \|^2$$

- If cluster centers are known…

$$\min\ f(S, M) = \min_{c_1, \ldots, c_k} \sum_{i=1}^{n} \min\{ \|s_i - c_1\|^2, \ldots, \|s_i - c_k\|^2 \}$$

- If the points in cluster are fixed…

$$\arg\min_{c_j}\ f(S, M) = \frac{1}{|M_j|} \sum_{i=1}^{|M_j|} s_i^{(j)}$$

K-means

- Algorithm
  1. Choose k cluster centers randomly generated in a domain containing all the points
  2. Assign each point to the closest cluster center
  3. Recompute the cluster centers using the current cluster memberships
  4. If a convergence criterion is met, stop; Otherwise go to step 2.

K-means

Rutgers
Rutgers Business School
Newark and New Brunswick

- Set the assignment matrix: $X = [x_{ij}] \in R^{n \times k}$; if s_i is assigned to M_j then $x_{ij} = 1$, otherwise $x_{ij} = 0$
- The cluster center of the cluster M_j as the mean of all the points in the cluster:

$$c_j = \frac{\sum_{l=1}^{n} x_{lj} s_l}{\sum_{l=1}^{n} x_{lj}}$$

K-means

RUTGERS
Rutgers Business School
Newark and New Brunswick

$$\min_{x_{ij}} \quad \sum_{j=1}^{k} \sum_{i=1}^{n} x_{ij} \left\| s_i - \frac{\sum_{l=1}^{n} x_{lj} s_l}{\sum_{l=1}^{n} x_{lj}} \right\|^2$$

$$s.t. \quad \sum_{j=1}^{k} x_{ij} = 1, \quad i = 1, \ldots n$$ each point is assigned to only one cluster

$$\sum_{i=1}^{n} x_{ij} \geq 1, \quad j = 1, \ldots k$$ each cluster contains at least one point

$$x_{ij} \in \{0, 1\}, \quad i = 1, \ldots n, \; j = 1, \ldots k$$

0-1 SDP

Rutgers Business School
Newark and New Brunswick

SDP:

$$\begin{aligned} \min \quad & \mathrm{Tr}(WZ) \\ s.t. \quad & \mathrm{Tr}(B_i Z) = b_i, \quad i = 1, \ldots m \\ & Z \succeq 0 \end{aligned}$$

If we replace SD constraint by the requirement that $Z^2 = Z$

0-1 SDP:

$$\begin{aligned} \min \quad & \mathrm{Tr}(WZ) \\ s.t. \quad & \mathrm{Tr}(B_i Z) = b_i, \quad i = 1, \ldots m \\ & Z^2 = Z,\ Z = Z^T \end{aligned}$$

Reformulation

$$\sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}\|s_i - c_j\|^2 = \sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}\left[\|s_i\|^2 + \|c_j\|^2 - 2s_i \cdot c_j\right]$$

$$= \sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}\|s_i\|^2 + \sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}\|c_j\|^2 - \sum_{j=1}^{k}\sum_{i=1}^{n} 2x_{ij}s_i \cdot c_j$$

$$= \sum_{i=1}^{n} \|s_i\|^2 \left(\sum_{j=1}^{k} x_{ij}\right) + \sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}\frac{\left\| \sum_{l=1}^{n} x_{lj}s_l \right\|^2}{(\sum_{l=1}^{n} x_{lj})^2} - \sum_{j=1}^{k} 2\frac{\left(\sum_{l=1}^{n} x_{lj}s_l\right)^2}{\sum_{l=1}^{n} x_{lj}}$$

$$= \sum_{i=1}^{n} \|s_i\|^2 - \sum_{j=1}^{k} \frac{\left\| \sum_{i=1}^{n} x_{ij}s_i \right\|^2}{\sum_{i=1}^{n} x_{ij}}$$

Reformulation

RUTGERS
Rutgers Business School
Newark and New Brunswick

Let $W_S \in R^{n \times d}$ denote the matrix whose i-th row is s_i^T.

$$\sum_{j=1}^{k} \sum_{i=1}^{n} x_{ij} \|s_i - c_j\|^2 = \sum_{i=1}^{n} \|s_i\|^2 - \sum_{j=1}^{k} \frac{\left\| \sum_{i=1}^{n} x_{ij} s_i \right\|^2}{\sum_{i=1}^{n} x_{ij}}$$

$$= \mathrm{Tr}(W_S W_S^T) - \sum_{j=1}^{k} \frac{\left\| \sum_{i=1}^{n} x_{ij} s_i \right\|^2}{\sum_{i=1}^{n} x_{ij}}$$

Reformulation

Since X is an assignment matrix, we have

$$X^T X = \mathrm{diag}\left(\sum_{i=1}^{n} x_{i1}^2, \ldots, \sum_{i=1}^{n} x_{ik}^2\right) = \mathrm{diag}\left(\sum_{i=1}^{n} x_{i1}, \ldots, \sum_{i=1}^{n} x_{ik}\right)$$

Let $Z = X(X^T X)^{-1} X^T$

Z is a projection matrix that $Z^2 = Z$ with nonnegative elements

e.g.

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad Z = \begin{pmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \end{pmatrix}, \quad Z^2 = Z, \quad ZX = X, \quad \mathrm{Tr}(Z) = k$$

eigenvalues are either 0 or 1

Reformulation

$$\mathrm{Tr}(W_S W_S^T) - \sum_{j=1}^{k} \frac{\left\| \sum_{i=1}^{n} x_{ij} s_i \right\|^2}{\sum_{i=1}^{n} x_{ij}}$$

$$\Rightarrow \mathrm{Tr}[W_S W_S^T (I - Z)] = \mathrm{Tr}(W_S^T W_S) - \mathrm{Tr}(W_S^T W_S Z)$$

For any integer m, let e^m be the all one vector in R^m, we have

$$\sum_{j=1}^{k} x_{ij} = 1 \quad \Rightarrow \quad Xe^k = e^n$$

It follows $Ze^n = ZXe^k = Xe^k = e^n$

0-1 SDP Model

$$\min_{x_{ij}} \quad \mathrm{Tr}[W_S W_S^T (I - Z)]$$

$$s.t. \quad Ze = e,\ \mathrm{Tr}(Z) = k$$

$$Z \geq 0,\ Z = Z^T,\ Z^2 = Z$$

- THEOREM 2.2:

Solving the 0-1 SDP problem is equivalent to finding a global solution of the original K-means integer programming problem.

0-1 SDP Model

$$\begin{aligned} \min_{x_{ij}} \quad & \mathrm{Tr}[W_S W_S^T (I - Z)] \\ s.t. \quad & Ze = e,\ \mathrm{Tr}(Z) = k \\ & Z \geq 0,\ Z = Z^T,\ Z^2 = Z \end{aligned}$$

- Note that for a given data set S, the trace $\mathrm{Tr}(W_S W_S^T)$ becomes a fixed quantity, therefore we can solve the MSSC model via following problem

$$\begin{aligned} \max_{x_{ij}} \quad & \mathrm{Tr}(W_S W_S^T Z) \\ s.t. \quad & Ze = e,\ \mathrm{Tr}(Z) = k \\ & Z \geq 0,\ Z = Z^T,\ Z^2 = Z \end{aligned}$$

Algorithms for 0-1 SDP

- Algorithm
  1. Choose a relaxation model
  2. Solve the relaxed problem for an approximate solution
  3. Use a rounding procedure to extract a feasible solution from the approximate solution.

Relaxation

$$\begin{aligned} \min \quad & \mathrm{Tr}[W_S W_S^T (I - Z)] \\ s.t. \quad & Ze = e,\ \mathrm{Tr}(Z) = k \\ & Z \geq 0,\ Z = Z^T,\ Z^2 = Z \end{aligned}$$

Replacing $Z^2 = Z$ by $I \succeq Z \succeq 0$, $I \succeq Z$ can be waived

$$\begin{aligned} \min \quad & \mathrm{Tr}(W(I - Z)) \\ s.t. \quad & Ze = e,\ \mathrm{Tr}(Z) = k \\ & Z \geq 0,\ Z \succeq 0 \end{aligned}$$

Could be solved by interior-point methods

Relaxation

$$
\begin{aligned}
\min \quad & \mathrm{Tr}[W_S W_S^T (I - Z)] \\
s.t. \quad & Ze = e, \ \mathrm{Tr}(Z) = k \\
& Z \geq 0, \ Z = Z^T, \ Z^2 = Z
\end{aligned}
$$

Replacing $Z^2 = Z$ and nonnegative requirement by $I \succeq Z \succeq 0$

$$
\begin{aligned}
\min \quad & \mathrm{Tr}(W(I - Z)) \\
s.t. \quad & Ze = e, \ \mathrm{Tr}(Z) = k \\
& I \succeq Z \succeq 0
\end{aligned}
$$

We will discuss how to solve it, denoted by P.

Relaxation

Rutgers Business School
Newark and New Brunswick

If Z is a feasible solution for P, then we have

$$\frac{1}{\sqrt{n}} Ze = \frac{1}{\sqrt{n}} e,$$

which implies $e/\sqrt{n}$ is an eigenvector of Z corresponding to its largest eigenvalue 1. For any feasible solution of P, define

$$Z_1 = Z - \frac{1}{n} ee^T$$

It's easy to see that

$$Z_1 = (I - \frac{1}{n} ee^T) Z = (I - \frac{1}{n} ee^T) Z (I - \frac{1}{n} ee^T)$$

Z1 is the projection of the matrix Z onto the null space of e

$$\mathrm{Tr}(Z_1) = \mathrm{Tr}(Z) - 1 = k - 1$$

Relaxation

Rutgers
Rutgers Business School
Newark and New Brunswick

If W1 is the projection of the matrix W onto null space of e

$$W_1 = (I - \frac{1}{n}ee^T)W(I - \frac{1}{n}ee^T)$$

Here we could reduce P to

$$\begin{aligned} \min \quad & \mathrm{Tr}(W_1(I - Z_1)) \\ s.t. \quad & \mathrm{Tr}(Z_1) = k - 1 \\ & I \succeq Z_1 \succeq 0 \end{aligned}$$

Let λ be the eigenvalues of the W1 listed in decreasing order. The optimal solution of this problem can be achieved if and only if

$$\mathrm{Tr}(W_1 Z_1) = \sum_{i=1}^{k-1} \lambda_i$$

Relaxation Algorithm

Rutgers Business School
Newark and New Brunswick

1. Calculate the projection W1 via $W_1 = (I - \frac{1}{n}ee^T)W(I - \frac{1}{n}ee^T)$
2. Compute the first k-1 largest eigenvalues of W1 and their corresponding eigenvectors v.
3. Set $Z = \frac{1}{n}ee^T + \sum_{i=1}^{k-1} v^i {v^i}^T$

Theorem 3.1

If Z* be the global optimal solution of 0-1 SDP, then

$$\mathrm{Tr}(W(I - Z^*)) \geq \mathrm{Tr}(W) - \frac{1}{n}e^T We - \sum_{i=1}^{k-1} \lambda_i$$

RUTGERS
Rutgers Business School
Newark and New Brunswick

Thanks