Low Rank Representation on Product Grassmann Manifolds for Multi-view Subspace Clustering

Jipeng Guo

Beijing University of Technology

2021.1
Content

- Subspace clustering
- Related Works
  1) Low Rank Representation (LRR)
  2) Low Rank Representation on Grassmann Manifolds (G-LRR)
  3) Low Rank Representation on Product Grassmann Manifolds (PG-LRR)
- Low Rank Representation with Matrix Factorization on Product Grassmann Manifolds (PG-MFLRR)
Subspace clustering

- Given sufficient data samples drawn from multiple low dimensional subspaces:

  the **goal** is to group a set of data samples into several clusters, which clusters corresponds to the independent subspaces.
Spectral Clustering (two steps):

- Graph construction: construct a graph (i.e., affinity matrix) to measure the similarities between data samples;

- Spectral clustering algorithm group the data samples into multiple clusters.
The basic assumption: data points are sampled from a union of \(k\) independent subspaces.

- Graph construction by low rank representation

\[
\min_Z \ \text{rank}(Z) \ \text{ s.t. } X = XZ
\]

\[
\min_Z \ \|Z\|_* \ \text{ s.t. } X = XZ
\]

\[
\min_{Z,E} \ \|Z\|_* + \alpha \|E\|_{2,1} \ \text{ s.t. } X = XZ + E
\]

- Graph matrix \(W = \frac{|Z|+|Z|}{2}\).
the high-dimension data in general lie in or close to a low dimensional manifold. Thus, extending the LRR based methods on the Grassmann manifold for high-dimension data clustering with the non-Euclidean geometry.

Related Works: G-LRR
Grassmann manifolds, denoted by $\mathcal{G}(p, d)$ is the space of all $p$-dimensional linear subspaces of $R^d (0 \leq p \leq d)$.

Grassmann manifolds can be embedded into the space of symmetric matrices $Sym(d)$ as

$$\pi: \mathcal{G}(p, d) \to Sym(d), \quad \pi(X) = XX^T.$$

Replace the distance on Grassmann manifolds with the following distance defined on the symmetric matrix space

$$d_{\mathcal{G}}^2 (X, Y) = \frac{1}{2} \|\pi(X) - \pi(Y)\|_F^2.$$
For a given sample set $\mathbf{X} = \{X_1, X_2, \cdots, X_n\}$ where $X_i \in \mathcal{G}(p, d)$, the G-LRR is formulated as:

$$
\min_Z \sum_{i=1}^{n} \left\| X_i \ominus (\cup_{j=1}^{n} z_{ij} \odot X_j) \right\|_{\mathcal{G}} + \alpha \|Z\|_* 
$$

where abstract symbols $\ominus, \cup_{j=1}^{n}, \odot$ denote the "linear" operations to be defined on manifolds, i.e., addition, subtraction and scalar multiplication. $\left\| X_i \ominus (\cup_{j=1}^{n} z_{ij} \odot X_j) \right\|_{\mathcal{G}}$ with operator $\ominus$ representing the product manifold distance between $X_i$ and its reconstruction $\cup_{j=1}^{n} z_{ij} \odot X_j$.
Related Works: PG-LRR

Given $V$ Grassmann manifolds with dimensions $p_1, \cdots, p_V$ respectively, the Product Grassmann manifolds (PGM) (denoted by $\mathcal{PG}(d; p_1, \cdots, p_V)$) is defined as $\mathcal{G}(p_1, d) \times \cdots \times \mathcal{G}(p_V, d)$.

Then, a point embedded in PGM is a set of Grassmann points, denoted by $[X] = \{X^1, \cdots X^V\}$ where $X^i \in \mathcal{G}(p_i, d)$. 
A valid distance on PGM can be induced from the individual distance on each Grassmann manifold as follows

\[ d^2_{\mathcal{P}_{\mathcal{G}}}(X, Y) = \sum_{\nu=1}^{V} d^2_{\mathcal{G}}(X^\nu, Y^\nu) \]
Related Works: PG-LRR

\[ \mathcal{X} = \{[X_1], \ldots, [X_n]\} \text{ be a set of given PGM samples,} \]

where \([X_i] = \{X_i^1, \ldots, X_i^V\} \in \mathcal{PG}(d; p_1, \ldots, p_V)\) with the basic matrix \(X_i^v \in \mathcal{G}(p_v, d)\). Then, PG-LRR is formulated as

\[
\min_Z \sum_{i=1}^{n} \| [X_i] \ominus (\cup_{j=1}^{n} Z_{ij} \odot [X_j]) \|_{\mathcal{PG}} + \alpha \|Z\|_*
\]

Remark: the objective of PG-LRR is convex of \(Z\).

Motivation: LRR, G-LRR, and PG-LRR directly employ convex nuclear norm $\|Z\|_*$ to approximate low rank constraint $\text{rank}(Z)$, which may be a biased estimation of the rank. Despite the elegant theory of the convex relaxation of $\text{rank}(Z)$, it has two major weaknesses:

1) over-relaxation of rank components leads to the results which can be far from the true underlying ones;

2) the singular value decomposition (SVD) of matrix has high complexity in computation.
PG-MFLRR

\[
min_Z \sum_{i=1}^{n} \| [X_i] \ominus (\bigcup_{j=1}^{n} Z_{ij} \odot [X_j]) \|_{\mathcal{PG}} + \alpha \|M\|_*
\]

s.t. \( Z = UMV^T, U^T U = V^T V = I_k \),

Where the fixed-rank \( k \ll n \) is a reasonable assumption which provides the approximate rather than random upper bound for true rank of \( Z \), resulting in a more accurate representation.
Theorem. A global minimum of non-convex model PG-MFLRR can always be obtained.

Proof: we can construct the optimal solution skillfully according to the optimum of convex model PG-LRR.

Computational complexity analysis

PG-LRR: $\mathcal{O}(t(2n^3))$
PG-MFLRR: $\mathcal{O}(t(n^3 + 3n^2k))$
Experiments: Multi-video clustering

<table>
<thead>
<tr>
<th></th>
<th>G-LRR</th>
<th>MVGL</th>
<th>MCGC</th>
<th>SM2SC</th>
<th>LCRSR</th>
<th>PG-LRR</th>
<th>PG-MFLRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>0.4541</td>
<td>0.1269</td>
<td>0.1429</td>
<td>0.1463</td>
<td>0.3766</td>
<td>0.4957</td>
<td>0.5098</td>
</tr>
<tr>
<td>NMI</td>
<td>0.5421</td>
<td>0.0492</td>
<td>0.0649</td>
<td>0.0628</td>
<td>0.2397</td>
<td>0.6250</td>
<td>0.6421</td>
</tr>
<tr>
<td>F-score</td>
<td>0.1296</td>
<td>0.1298</td>
<td>0.1445</td>
<td>0.0782</td>
<td>0.2217</td>
<td>0.5102</td>
<td>0.5317</td>
</tr>
</tbody>
</table>

Table 1. Clustering results on **ACT4** video database
Experiments: Multi-video clustering

<table>
<thead>
<tr>
<th></th>
<th>G-LRR</th>
<th>MVGL</th>
<th>MCGC</th>
<th>SM2SC</th>
<th>LCRSR</th>
<th>PG-LRR</th>
<th>PG-MFLRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>0.2904</td>
<td>0.2775</td>
<td>0.2679</td>
<td>0.1100</td>
<td>0.2700</td>
<td>0.2969</td>
<td>0.3560</td>
</tr>
<tr>
<td>NMI</td>
<td>0.2202</td>
<td>0.2024</td>
<td>0.1972</td>
<td>0.0042</td>
<td>0.2078</td>
<td>0.3525</td>
<td>0.3681</td>
</tr>
<tr>
<td>F-score</td>
<td>0.2976</td>
<td>0.2932</td>
<td>0.2710</td>
<td>0.1798</td>
<td>0.0641</td>
<td>0.3017</td>
<td>0.3978</td>
</tr>
</tbody>
</table>

Table 1. Clustering results on **NUCLA** video database
Experiments: Multi-video clustering

<table>
<thead>
<tr>
<th></th>
<th>G-LRR</th>
<th>MVGL</th>
<th>MCGC</th>
<th>SM2SC</th>
<th>LCRSR</th>
<th>PG-LRR</th>
<th>PG-MFLRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>0.4196</td>
<td>0.4041</td>
<td>0.4071</td>
<td>0.3835</td>
<td>0.3890</td>
<td>0.4240</td>
<td><strong>0.4945</strong></td>
</tr>
<tr>
<td>NMI</td>
<td>0.4669</td>
<td>0.4831</td>
<td>0.4448</td>
<td>0.4095</td>
<td>0.3740</td>
<td>0.4773</td>
<td><strong>0.5014</strong></td>
</tr>
<tr>
<td>F-score</td>
<td>0.4187</td>
<td>0.4289</td>
<td>0.2710</td>
<td>0.2808</td>
<td>0.3889</td>
<td>0.4268</td>
<td><strong>0.4912</strong></td>
</tr>
</tbody>
</table>

Table 1. Clustering results on IXMAS video database
Experiments: Multi-video clustering

<table>
<thead>
<tr>
<th></th>
<th>G-LRR</th>
<th>MVGL</th>
<th>MCGC</th>
<th>SM2SC</th>
<th>LCRSR</th>
<th>PG-LRR</th>
<th>PG-MFLRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>0.7802</td>
<td>0.7802</td>
<td>0.9396</td>
<td>0.7637</td>
<td>0.8381</td>
<td>0.8022</td>
<td>1.0000</td>
</tr>
<tr>
<td>NMI</td>
<td>0.6870</td>
<td>0.6870</td>
<td>0.8436</td>
<td>0.5275</td>
<td>0.7237</td>
<td>0.6075</td>
<td>1.0000</td>
</tr>
<tr>
<td>F-score</td>
<td>0.7059</td>
<td>0.7913</td>
<td>0.9393</td>
<td>0.6399</td>
<td>0.7734</td>
<td>0.8014</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Table 1. Clustering results on DTHC video database
Thanks!