Scalable Sparse Subspace Clustering by Orthogonal Matching Pursuit

Chong You Daniel P. Robinson René Vidal

Johns Hopkins University, Baltimore, MD, 21218, USA

Center for IMAGING SCIENCE, JOHNS HOPKINS UNIVERSITY
Vision Lab @ JHU
http://www.vision.jhu.edu

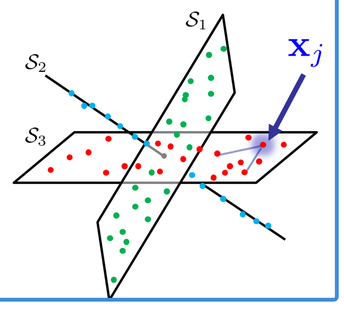
IEEE 2016 Conference on Computer Vision and Pattern Recognition
CVPR 2016

Motivation

- Vision datasets often contain multiple classes, each lying in a low-dimensional subspace
- **Subspace clustering**: cluster data that lie in a union of subspaces

Spectral Subspace Clustering

- Approach
  - Step 1: build data affinity
  - Step 2: apply spectral clustering
- Challenge: distance based affinity fails at the intersection of subspaces
- Solution: learn affinity by data self representation, i.e., $\mathbf{x}_j = X\mathbf{c}_j$, where $X = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$



Sparse Subspace Clustering (SSC)

- Learn affinity by finding the sparsest data self representation

$$\min_{\mathbf{c}_j} \|\mathbf{c}_j\|_0 \quad \text{s.t.} \quad \mathbf{x}_j = X\mathbf{c}_j, \quad c_{jj} = 0$$

Prior Method

- Basis Pursuit (BP):
  - Replace $\|\mathbf{c}_j\|_0$ with $\|\mathbf{c}_j\|_1$
- Properties:

✓ Guaranteed correct connections under broad conditions

✗ Not scalable: solved by the CVX software or ADMM algorithm

Contribution

- Orthogonal Matching Pursuit (OMP):
  - Choose one point at a time
- Properties:

✓ Guaranteed correct connections under broad conditions

✓ Scalable: performance is verified on 100,000 data points

Guaranteed correct connections

Theorem

Suppose that $\mathbf{x}_j \in \mathcal{S}_\ell$. Then, $\mathbf{c}_j$ gives correct connections if

$$\mu(W_j^\ell, X^{-\ell}) < r^\ell$$

- μ captures the similarity between $\mathcal{S}_\ell$ and all other subspaces
- r captures the distribution of points in subspace $\mathcal{S}_\ell$
- W_j^ℓ is the dual points/residual points for SSC-BP/SSC-OMP

Theorem

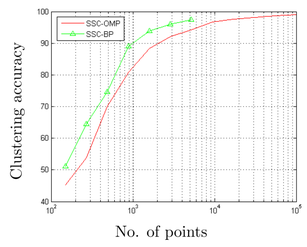
Randomly generate n subspaces of dimension d in ambient dimension D, and $\rho d + 1$ points from each subspace. $\{\mathbf{c}_j\}_{j=1}^N$ gives correct connections with high probability if

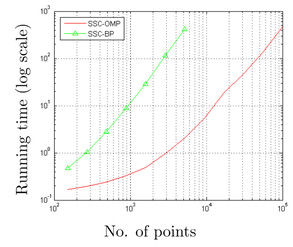
$$\frac{d}{D} < \frac{c^2(\rho)\log\rho}{12\log N}$$

- Same condition for SSC-BP/SSC-OMP, with different probability

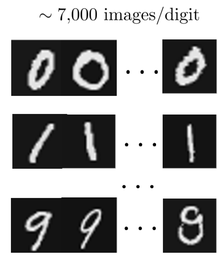
Experiments on Synthetic, Face Image and Digit Image Data

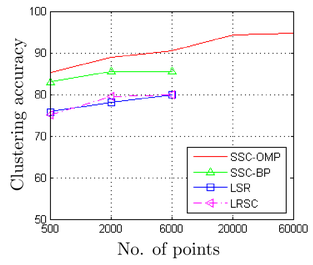
- Synthetic data: draw 5 subspaces of dimension 6 in ambient dimension 9; draw equal number of points from each subspace
- Clustering accuracy: SSC-OMP is slightly outperformed by SSC-BP, and the difference decreases as number of points increases
- Running time: SSC-OMP is orders of magnitude faster than SSC-BP, and is able to handle up to 100,000 points efficiently

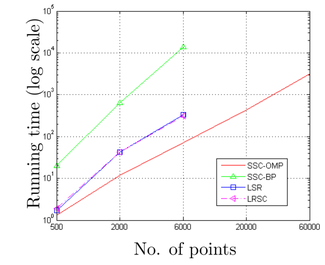




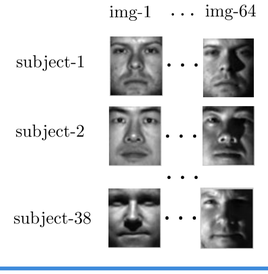
- MNIST handwritten digit database
- First time to test on 60,000 images
- Clustering accuracy: SSC-OMP obtains the best performance
- Running time: SSC-OMP can handle more points than other methods

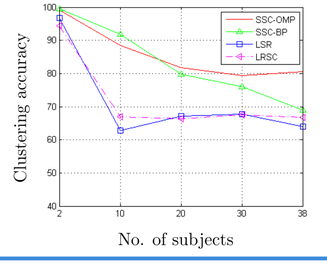


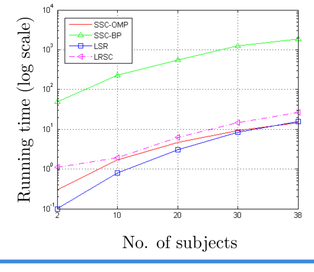




- Extended Yale B face database
- First time to test on all 38 subjects
- Clustering accuracy: SSC-OMP and SSC-BP achieves state of the art
- Running time: SSC-OMP is $>$ 100 times faster than SSC-BP







[1] E. Elhamifar and R. Vidal., Sparse Subspace Clustering, In *IEEE Conf. in Computer Vision and Pattern Recognition*, 2009.
[2] M. Soltanolkotabi and E.J. Candes., A Geometric Analysis of Subspace Clustering with Outlier, In *Annals of Statistics*, 2013.
[3] E. Dyer et al., Greedy Feature Selection for Subspace Clustering, In *Journal of Machine Learning Research*, 2014.

This work was supported by the grants NSF-IIS 1447822.