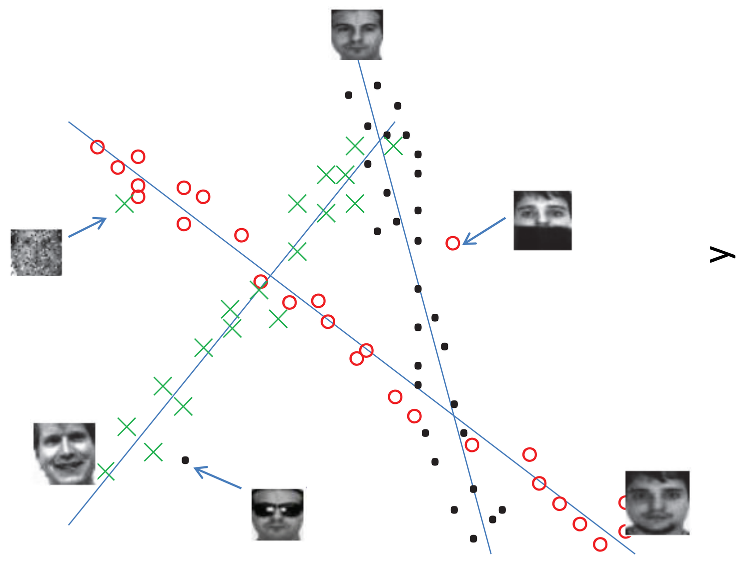
Correntropy Induced L2 Graph for Robust Subspace Clustering

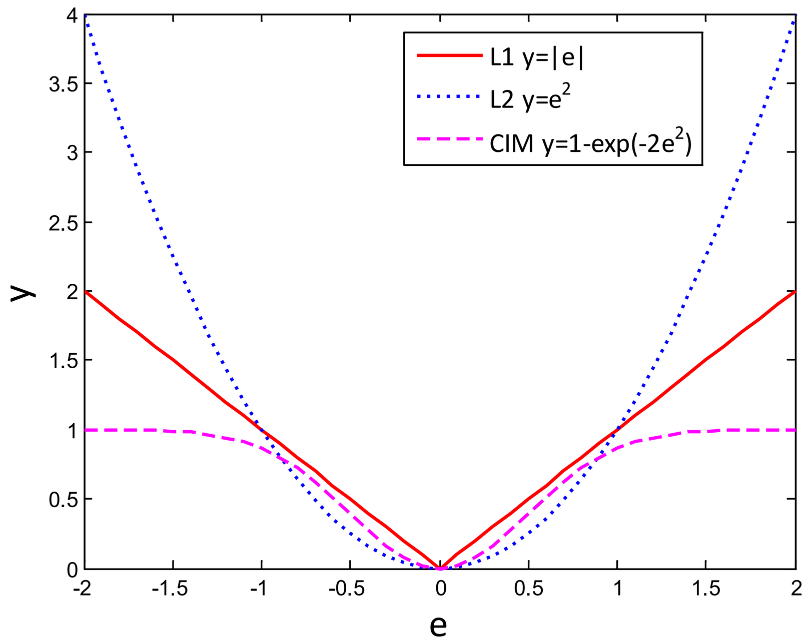
Canyi Lu, Jinhui Tang, Min Lin, Liang Lin, Shuicheng Yan, Zhouchen Lin
National University of Singapore

Robust Subspace Clustering

Given a set of data vectors $\mathbf{X} = [\mathbf{X}_1, \cdots, \mathbf{X}_k] = [\mathbf{x}_1, \cdots, \mathbf{x}_n] \in \mathbb{R}^{d\times n}$ drawn from a union of k subspaces $\{\mathcal{S}_i\}_{i=1}^k$. Let $\mathbf{X}_i$ be a collection of n_i data vectors drawn from the subspace $\mathcal{S}_i$, $n = \sum_{i=1}^k n_i$. The task is to group the data according to the underlying subspaces they are drawn from. We aims to propose a robust method to handle the problem with non-Gaussian and impulsive noises.



Corruptions deviate the data from the underlying subspaces.



Comparison of different loss functions.

Related Works

- Spectral Clustering (SC) is used as the framework for subspace clustering.
- The main challenge by using SC is to define a "good" affinity matrix (or graph) $\mathbf{Z} \in \mathbb{R}^{n\times n}$. Each entry z_{ij} measures the similarity between $\mathbf{x}_i$ and $\mathbf{x}_j$.

For a given data matrix $\mathbf{X} \in \mathbb{R}^{d\times n}$, many previous methods fall into the following general model :

$$\min_{\mathbf{Z}} \mathcal{J}(\mathbf{Z}) = \lambda\mathcal{R}(\mathbf{Z}) + \mathcal{L}(\mathbf{E}) \quad \text{s.t.} \quad \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E},$$

where $\mathcal{R}(\mathbf{Z})$ is the regularizer, and $\mathcal{L}(\mathbf{E})$ is the loss function.

L1-graph (S. Yan, SDM 2009) or **SSC (Sparse Subspace Clustering)** (Elhamifar, CVPR 2009):

$$\mathcal{R}(\mathbf{Z}) = ||\mathbf{Z}||_1, \quad \text{and} \quad \mathcal{L}(\mathbf{E}) = ||\mathbf{E}||_F^2, \quad \text{or} \quad ||\mathbf{E}||_1.$$

LRR (Low-Rank Representation) (G. Liu et al., ICML 2010, TPAMI 2012):

$$\mathcal{R}(\mathbf{Z}) = ||\mathbf{Z}||_*, \quad \text{and} \quad \mathcal{L}(\mathbf{E}) = ||\mathbf{E}||_{21}.$$

LSR (Least Squares Regression) (C. Lu and S. Yan, ECCV 2012):

$$\mathcal{R}(\mathbf{Z}) = ||\mathbf{Z}||_F^2, \quad \text{and} \quad \mathcal{L}(\mathbf{E}) = ||\mathbf{E}||_F^2.$$

Different from the previous methods which focus on the regularization term $\mathcal{R}(\mathbf{Z})$, this work focuses on the error term $\mathcal{L}(\mathbf{Z})$ for robust subspace learning.

- Frobenius norm is optimal for Gaussian noise.
- L1-norm can handle sparse errors with large magnitude.
- L21-norm is able to remove the outlier samples, but it is sensitive to the outlier features.

Correntropy Induced L2 Graph for Robust Subspace Clustering

Correntropy Induced Metric

The Correntropy Induced Metric (CIM) of two vectors $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$ is formally defined as

$$\text{CIM}(\mathbf{x}, \mathbf{y}) = 1 - \frac{1}{n}\sum_{i=1}^n k_\sigma(x_i - y_i),$$

where $k_\sigma(e) = \exp\left(-e^2/2\sigma^2\right)$.

- The mean squared error (L2-norm) is a global metric which increases quadratically for large errors.
- L1-norm is more robust, but is also sensitive when the error is very large.
- CIM is a local metric which is close to $L1$-norm when the errors are relatively small. For large errors, the value of CIM is close to 1. Note that the large errors are usually caused by outliers, but their effect on CIM is limited. Therefore CIM will be more robust to the non-Gaussian noises.

Correntropy Induced L2 Graph (CIL2)

We use the correntropy to model the reconstruction error, leading to the **Correntropy Induced L2 (CIL2)** graph as follows:

$$\min_{Z} \sum_{i,j}(1 - k_\sigma(e_{ij})) + \lambda||\mathbf{Z}||_F^2 \quad \text{s.t.} \quad \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}.$$

The above problem can be solved by half-quadratic analysis, which alternately updates ($i = 1, \cdots, d$, $j = 1\cdots, n$)

$$s_{ij}^{t+1} = \frac{1}{(\sigma^t)^2}\exp\left(-(e_{ij}^t)^2/2(\sigma^t)^2\right),$$

$$\mathbf{z}_j^{t+1} = \arg\min_{\mathbf{z}_j}(\mathbf{x}_j - \mathbf{X}\mathbf{z}_j)^T\text{Diag}(\mathbf{s}_j^{t+1})(\mathbf{x}_j - \mathbf{X}\mathbf{z}_j) + \lambda||\mathbf{z}_j||_2^2.$$

In each iteration, the kernel size σ is empirically updated by the average reconstruction error

$$(\sigma^{t+1})^2 = \frac{1}{2dn}||\mathbf{X} - \mathbf{X}\mathbf{Z}^t||_F^2.$$

Row based Correntropy Induced L2 Graph (rCIL2)

To handle the case that the rows/features are corrupted, e.g. the face images with sunglasses and scarf are outliers, we propose the **row based Correntropy Induced L2 (rCIL2)** graph as follows

$$\min_{\mathbf{Z}} \sum_{i=1}^d \left(1 - k_\sigma(||\mathbf{e}^i||_2)\right) + \lambda||\mathbf{Z}||_F^2 \quad \text{s.t.} \quad \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}.$$

The above problem can be solved by half-quadratic analysis.

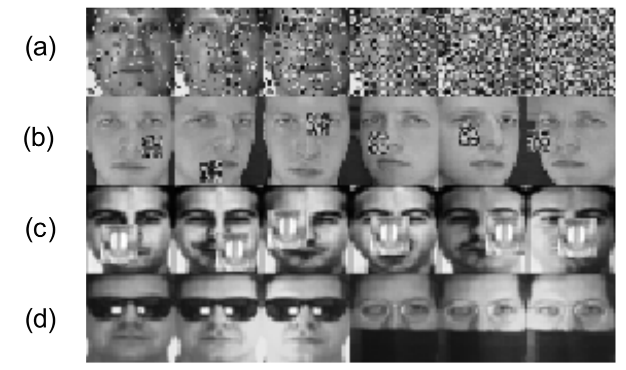
Experiments

Evaluation Metrics: the Accuracy and Normalized Mutual Information (NMI) metrics are used:

$$Accuracy = \frac{\sum_{i=1}^n \delta(y_i, \text{map}(p_i))}{n},$$

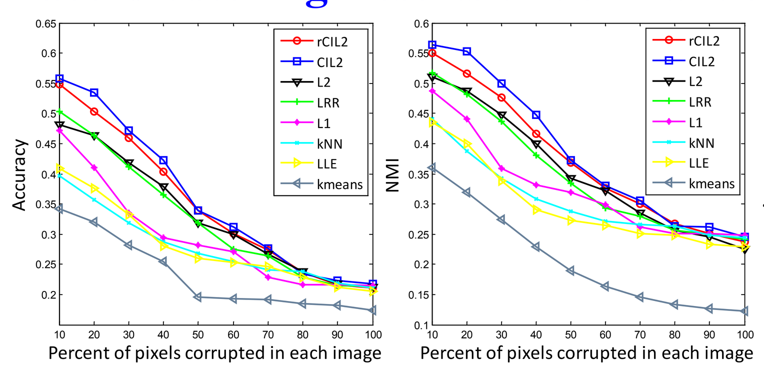
$$NMI(C, C') = \frac{MI(C, C')}{\max(H(C), H(C'))},$$

map(p_i): the permutation mapping function;
$MI(C, C')$: the mutual information metric;
$H(C)$: the entropy of the cluster set C.

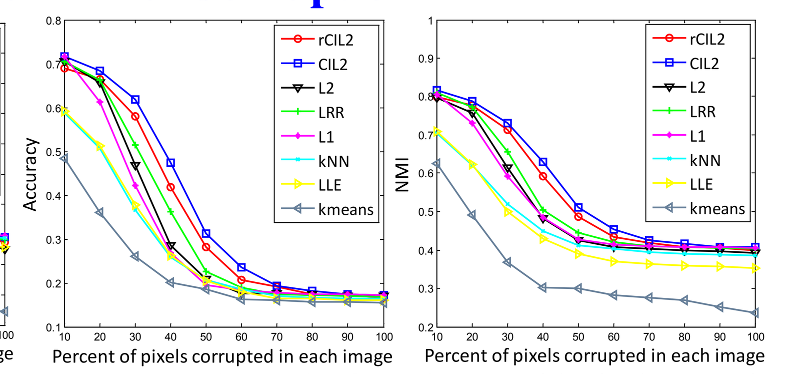


Face images with (a) random pixels corrupttion; (b) random block occlusion; (c) a monkey face; (d) sunglasses and scarf.

Face Clustering Results under Random Pixel Corruption

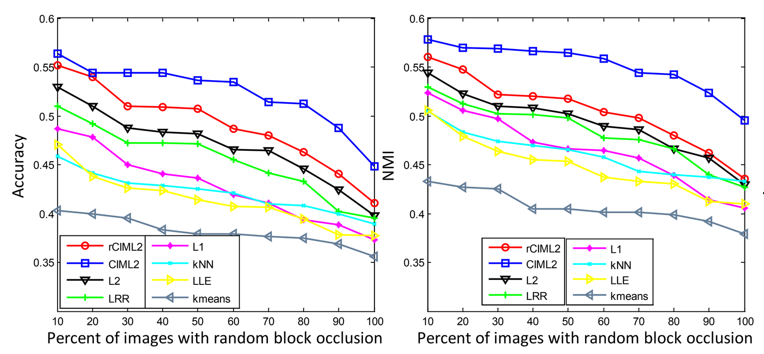


Results on the Yale dataset with pixel corruption.

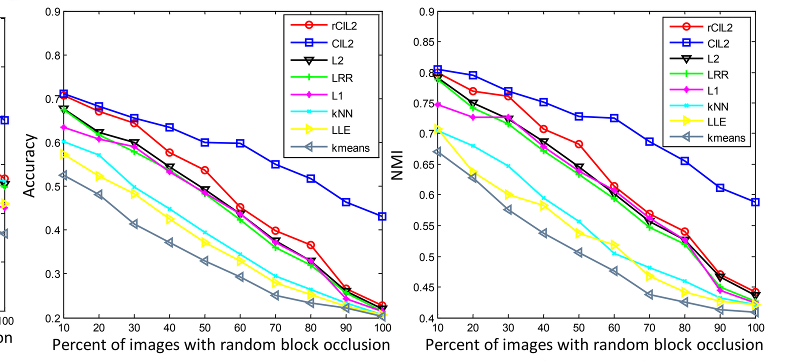


Results on the ORL dataset with pixel corruption.

Face Clustering Results under Contiguous Occlusion

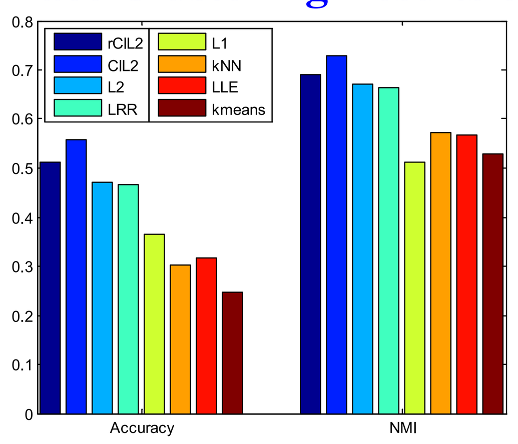


Results on the Yale dataset with block occlusion.



Results on the ORL dataset with block occlusion.

Face Clustering Results on Real-World Malicious Occlusion



Results on the AR dataset with an unrelated image occlusion.

Methods	Accuracy				NMI			
	Session 1		Session 2		Session 1		Session 2	
	Sunglasses	Scarf	Sunglasses	Scarf	Sunglasses	Scarf	Sunglasses	Scarf
rCIL2	**85.2**	**78.4**	**86.4**	**81.2**	**93.8**	**90.1**	**94.0**	**91.5**
CIL2	81.2	75.4	85.4	79.0	89.9	87.9	93.8	88.8
L2	78.2	71.6	80.0	72.6	86.3	83.8	90.7	83.8
LRR	77.2	72.2	79.6	74.6	86.6	84.7	90.7	84.7
L1	43.8	40.6	27.8	40.2	72.3	67.7	55.8	60.8
kNN	26.4	25.6	26.8	27.2	65.4	66.0	66.1	65.9
LLE	28.0	27.6	33.2	27.2	63.4	62.9	66.6	61.7
kmeans	30.0	29.4	30.8	29.8	65.4	63.8	65.3	65.5

Results on the AR dataset with sunglasses or scarfs.

Conclusions

- Different from most of previous works, which focus on the regularizer, we focus on the robustness of the loss function instead.
- We use the robust Correntropy Induced Metric (CIM) to measure the reconstruction error, which leads to the Correntropy Induced L2 (CIL2) graph. We further propose the row based Correntropy Induced L2 (rCIL2) graph to handle the whole rows/features outliers.
- Some other robust measure metrics may be used.