Classification and Clustering via Dictionary Learning with Structured Incoherence and Shared Features

Ignacio Ramirez, Pablo Sprechmann, and Guillermo Sapiro
University of Minnesota
Source separation

Applications of Sparse Models

Object recognition

Image segmentation
Sparse Models

Very successful when $D$ is learned from the data

Much better than using off-the-shelf dictionaries (DCT, Fourier, Wavelets)
Sparse Models

Very successful when $D$ is learned from the data

Much better than using off-the-shelf dictionaries (DCT, Fourier, Wavelets)
Sparse Models

Very successful when D is learned from the data

Much better than using off-the-shelf dictionaries (DCT, Fourier, Wavelets)
Sparse Models

Very successful when \( D \) is learned from the data

Much better than using off-the-shelf dictionaries (DCT, Fourier, Wavelets)

The union of subspaces: 
\[ A \] is sparse
Learning a Sparse Model

\[(A^*, D^*) = \arg \min_{A, D} \sum_{x \in C} \|x - Da\|^2 + \lambda \|a\|_1\]

- Atoms satisfy \(\|d_i\|_1 = 1\)
- Usual sparsity-inducing regularizers:
  - \(\ell_0\) “norm”: \(\psi(a_j) = \|a_j\|_0\)
  - \(\ell_1\) norm: \(\psi(a_j) = \|a_j\|_1\)

See also: SPAMS software.
Sparse Models for Supervised Classification

- Classes: \( \{C_1, C_2, \ldots, C_c\} \)
- Training: \( \{x^i_1, \ldots, x^i_{n_i}\} \subset C_i \)

Proposed Method
1. Learn (fit) a dictionary \( D_i \) to represent samples from class \( C_i \).
2. Use representation cost as discriminant function
\[
R(x, D_i) = \min_a ||x - D_i a||_2^2
\]
3. Assign sample to class with smallest \( R(x, D_i) \)

\[
\text{Class}(x) = \arg \min_i R(x, D_i)
\]

See also: [Mairal CVPR '08].
Sparse Models for Supervised Classification

- Classes: \( \{C_1, C_2, \ldots, C_c\} \)
- Training: \( \{x_1^i, \ldots, x_{n_i}^i\} \subset C_i \)

Proposed Method

1. Learn (fit) a dictionary \( D_i \) to represent samples from class \( C_i \).
2. Use representation cost as discriminant function
   \[
   R(x, D_i) = \min_a ||x - D_i a||^2_2 + \lambda ||a||_1
   \]
   Fitting term + Complexity term
3. Assign sample to class with smallest \( R(x, D_i) \)
   \[
   \text{Class}(x) = \arg \min_i R(x, D_i)
   \]

See also: [Mairal CVPR ‘08].
## Classification Results

### Classification Error Rate (%).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Mairal et al. NIPS '08</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNIST</td>
<td>1.0</td>
<td>3.4</td>
</tr>
<tr>
<td>USPS</td>
<td>3.5</td>
<td>4.4</td>
</tr>
</tbody>
</table>
Promoting Cross-Incoherence

\[
\min_{\{D_i, A_i\}_{i=1}^c} \sum_{i=1}^c \sum_{x \in C_i} \left\{ \|x - D_i a\|_2^2 + \lambda \|a\|_1 \right\}
\]
Promoting Cross-Incoherence

\[
\min_{\{D_i, A_i\}_{i=1}^{c}} \sum_{i=1}^{c} \sum_{x \in C_i} \left\{ \|x - D_i a\|^2_2 + \lambda \|a\|_1 \right\} + \eta \sum_{j \neq i} \|D_i^T D_j\|^2_F
\]

- More incoherence leads to better discriminative power

See also: [Tropp S.P. 2006] and [Eldar et al. TIT, Nov. 2009].
Promoting Cross-Incoherence

\[
\min_{\{D_i, A_i\}_{i=1}^c} \sum_{i=1}^c \sum_{x \in C_i} \left\{ \|x - D_i a\|_2^2 + \lambda \|a\|_1 \right\} + \eta \sum_{j \neq i} \|D_i^T D_j\|_F^2
\]

- More incoherence leads to better discriminative power
- Shared Features:

[Diagram showing two overlapping dictionaries labeled "3" and "5" with high coherence atoms]

See also: [Tropp S.P. 2006] and [Eldar et al. TIT, Nov. 2009].
Sparse Models for Clustering
Sparse Models for Clustering

1. learn global dict.
Sparse Models for Clustering

See also: L1 graph [Cheng TIP, Apr. 2010] and Subspace Clustering [Elhamifar CVPR ‘09].
Sparse Models for Clustering

See also: L1 graph [Cheng TIP, Apr. 2010] and Subspace Clustering [Elhamifar CVPR ‘09].
Sparse Models for Clustering

See also: L1 graph [Cheng TIP, Apr. 2010] and Subspace Clustering [Elhamifar CVPR ‘09].
Sparse Models for Clustering

- Energy Minimization Problem
- Lloyd’s type of algorithm for minimizing:

\[
\min_{C_i, \{D_i\}} \sum_{i=1}^{c} \sum_{x \in C_i} \min_{a} R(x, D_i) \leq \|x - D_i a\|_2^2 + \lambda \|a\|_1 + \eta \sum_{i \neq j} \left\| D_i^T D_j \right\|^2_F
\]
Object Detection

- Detection based on local descriptors.
- Learn Dictionaries for SIFT feature vectors.

See also: [Mairal CVPR ‘08, Yang CVPR ‘09].
Texture Segmentation

Dictionaries Learned on Image Patches

See also: [Peyre JMIIV, May 2008, Mairal et al. CVPR ‘08].
Extensions
Extensions

\[
\begin{align*}
\text{nonzero coefficient} & \\
\text{nonzero group} & \\
\text{zero} & \\
\end{align*}
\]

\[
\begin{array}{c}
x_1 \ x_2 \ x_3 \ \ldots \ x_n \\
\end{array}
\] = \[
\begin{array}{c}
D_1 \ \ldots \ D_c \\
\end{array}
\]

\[
\begin{array}{c}
x \\
\end{array}
\]
Extensions

nonzero coefficient
nonzero group
zero

\[ x_1 x_2 x_3 \ldots x_n = D_2 \ldots D_c x \]
Source Separation: Hierarchical models
Source Separation: Hierarchical models

See also: [Yuan and Lin 2006, Jenatton arXive, 2009].
Source Separation: Hierarchical models

See also: [Yuan and Lin 2006, Jenatton arXive, 2009].
Collaborative Source Separation

\[
\begin{align*}
\min_{A} & \quad \| X - DA \|_F^2 + \lambda_2 \sum_{g=1}^{c} \| A_g \|_F^2 + \lambda_1 \sum_{k=1}^{n} \| a_k \|_1 \\
\text{s.t.} & \quad A \in \mathbb{R}^{p \times n}
\end{align*}
\]

Collaborative Source Separation

\[
\min_{A \in \mathbb{R}^{p \times n}} \frac{1}{2} \left\| X - DA \right\|_F^2 + \lambda_2 \sum_{g=1}^{c} \left\| A_g \right\|_F + \lambda_1 \sum_{k=1}^{n} \left\| a_k \right\|_1
\]

Collaborative Source Separation Results

Recovery of (two) superimposed textures

<table>
<thead>
<tr>
<th>Potential Sources</th>
<th>D49</th>
<th>D84</th>
<th>D53</th>
<th>D52</th>
<th>D33</th>
<th>D3</th>
<th>D24</th>
<th>D6</th>
<th>mix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground Truth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recovered C-HiLasso</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Collaborative Source Separation Results

### Recovery of (two) superimposed textures

<table>
<thead>
<tr>
<th>D49</th>
<th>D84</th>
<th>D53</th>
<th>D52</th>
<th>D33</th>
<th>D3</th>
<th>D24</th>
<th>D6</th>
<th>mix</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="potential_sources_d49.png" alt="Image" /></td>
<td><img src="potential_sources_d84.png" alt="Image" /></td>
<td><img src="potential_sources_d53.png" alt="Image" /></td>
<td><img src="potential_sources_d52.png" alt="Image" /></td>
<td><img src="potential_sources_d33.png" alt="Image" /></td>
<td><img src="potential_sources_d3.png" alt="Image" /></td>
<td><img src="potential_sources_d24.png" alt="Image" /></td>
<td><img src="potential_sources_d6.png" alt="Image" /></td>
<td><img src="potential_sources_mix.png" alt="Image" /></td>
</tr>
</tbody>
</table>

#### Ground Truth

**Potential Sources**

**Recovered C-HiLasso**

### Recovery of superimposed numbers with missing information

<table>
<thead>
<tr>
<th>mixture 3+5</th>
<th>observed</th>
<th>recovered 3</th>
<th>recovered 5</th>
<th>Lasso</th>
<th>C-HiLasso</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="recovered_numbers_mixture_3_5.png" alt="Image" /></td>
<td><img src="recovered_numbers_observed.png" alt="Image" /></td>
<td><img src="recovered_numbers_recovered_3.png" alt="Image" /></td>
<td><img src="recovered_numbers_recovered_5.png" alt="Image" /></td>
<td><img src="recovered_numbers_lasso.png" alt="Image" /></td>
<td><img src="recovered_numbers_chilasso.png" alt="Image" /></td>
</tr>
</tbody>
</table>
Conclusions

- Framework for classification and clustering rich data via dictionary learning
- Simple metric derived from sparse modeling
- Inclusion of incoherence and shared features detection
- Extension to source separation
Conclusions

- Framework for classification and clustering rich data via dictionary learning
- Simple metric derived from sparse modeling
- Inclusion of incoherence and shared features detection
- Extension to source separation

Thank you!!