3+ Snippets on Learning Features: Material for Brainstorming

Guillermo Sapiro

Duke University
guillermo.sapiro@duke.edu
1.0: Dictionary Learning
   - 1.1: Sparse Modeling and Deep Learning?
     • Some food for though/discussion
   - 1.2: Multimodal Task Oriented Learning
   - 1.3: The Data are the Features

2.0: Helping the Data
   - 2.1: Another (Inverted) Deep Structure?

3.0: Discussion, Open Floor
1.0: Dictionary Learning
\[
\text{Min}_{D,A} \sum_{j=1}^{P} \left\| D\alpha_j - x_j \right\|_2^2 \quad \text{s.t.} \quad \forall j, \left\| \alpha_j \right\|_0^0 \leq L
\]

- State of the art in numerous image and audio challenges
1.1: Sparse Modeling and Deep Learning

Unstructured Lasso encoder architecture

RPCA/RNMF encoder architecture
Task Oriented and Multimodal
Query

CM-SSH:
sunset, tree, orange, old, abandoned, car, autumn, road, forest, fall, truck, rust, colourful, woods, antique, vehicle, halloween

MM-SparseHash:
clouds, sunset, sea, beach, sun, ocean, summer, sand, rocks, evening, holiday, peace, happy, dunes

ingland, italy, island, ship, italia, hawaii, interesting, cow, islands, elephants, maui

nature, sky, blue, water, clouds, red, sea, yellow, beach, california, winter, ocean, building, old, sand, sunrise, span, cloud, wall, coast, sepia, stone, eaves, mist, perspective, fence, school, fly, oregon, jump, monument, perfect, surf, alley

nature, sky, water, landscape, sunset, light, white, trees, color, reflection, black, animal, tree, sun, orange, winter, snow, beautiful, river, wildlife, photography, lake, bird, dark, forest, birds, ice, reflections, wood, flying, evening, outdoors, photographer, dusk

nature, sky, water, clouds, green, explore, sunset, people, sea, art, beach, ocean, asia, sand, rocks, airplane, aircraft, boats, flying, plane, rural, waves, flight, aviation, breathtaking, bush, thailand, vivid, twilight, glow, cliff, landscapes, airplanes
• Deep Learning
  – First train unsupervised
  – Then train supervised

• Sparse Modeling and Dictionary Learning
  – First train for generic data (“unsupervised”)
  – Then adapt to the data at hand (“supervised”)

• Excellent Supporting Theory
1.2: The Data are the Features

- Supporting Theory
2.0: Helping the Data, Learning the Features
Motivation
Motivation

(a) $\theta_{AB} = \frac{\pi}{2} = 1.57.$

(b) $T = \begin{bmatrix} 1.00 & 0 \\ 0 & 1.00 \end{bmatrix}$; $\theta_{AB} = 1.57.$

(c) $\theta_{AB} = \frac{\pi}{4} = 0.79.$

(d) $T = \begin{bmatrix} 0.50 & -0.21 \\ -0.21 & 0.91 \end{bmatrix}$; $\theta_{AB} = 1.57.$

(e) $\theta_{AB} = 0.79, \quad \theta_{AC} = 0.79, \quad \theta_{BC} = 1.05$
$\epsilon_A = 0.0141, \quad \epsilon_B = 0.0131, \quad \epsilon_C = 0.0148$
$|A|_* = 4.06, \quad |B|_* = 4.08, \quad |C|_* = 4.16.$

(f) $T = \begin{bmatrix} 0.39 & -0.16 & -0.16 \\ -0.13 & 0.90 & 0.11 \\ -0.23 & 0.11 & 0.57 \end{bmatrix}$; $\theta_{AB} = 1.51, \quad \theta_{AC} = 1.49, \quad \theta_{BC} = 1.57$
$\epsilon_A = 0.0091, \quad \epsilon_B = 0.0085, \quad \epsilon_C = 0.0114$
$|A|_* = 1.93, \quad |B|_* = 2.37, \quad |C|_* = 1.20.$
Nuclear Norm Formulation

\[
\arg\min_T \frac{1}{C} \sum_{c=1}^{C} \|TY_c\|_* - \lambda \|TY\|_*, \quad s.t. \|T\|_2 = 1.
\]

- **Theorem:** Positive
- **Theorem:** Zero for orthogonal subspaces
  - Not true for other popular norms
- Works on-line
- Works with **compressing** transform matrix
- Can learn class-specific transformations
- Building block of subspace clustering and classification
  - And beyond
Example

(a) $\theta_{AB} = \frac{\pi}{2} = 1.57$.  
(b) $T = \begin{bmatrix} 1.00 & 0 \\ 0 & 1.00 \end{bmatrix}$;  
\[ \theta_{AB} = 1.57. \]

(c) $\theta_{AB} = \frac{\pi}{4} = 0.79$.  
(d) $T = \begin{bmatrix} 0.50 & -0.21 \\ -0.21 & 0.91 \end{bmatrix}$;  
\[ \theta_{AB} = 1.57. \]

(e) $[\theta_{AB} = 0.79, \quad \theta_{AC} = 0.79, \quad \theta_{BC} = 1.05]$ \[ \epsilon_A = 0.0141, \quad \epsilon_B = 0.0131, \quad \epsilon_C = 0.0148 \] \[ |A|_* = 4.06, \quad |B|_* = 4.08, \quad |C|_* = 4.16. \]

(f) $T = \begin{bmatrix} 0.39 & -0.16 & -0.16 \\ -0.13 & 0.90 & 0.11 \\ -0.23 & 0.11 & 0.57 \end{bmatrix}$;  
\[ \theta_{AB} = 1.51, \quad \theta_{AC} = 1.49, \quad \theta_{BC} = 1.57 \] \[ \epsilon_A = 0.0091, \quad \epsilon_B = 0.0085, \quad \epsilon_C = 0.0114 \] \[ |A|_* = 1.93, \quad |B|_* = 2.37, \quad |C|_* = 1.20. \]
Example: Faces

- YaleB: 38 subjects and 64 lighting conditions
- CMU PIE: 68 subjects, 13 positions, 21 lighting conditions
Faces Clustering: 9 faces

(a) Original subspace angles.
(b) Transformed subspace angles.
(c) Subspace nuclear norm.

(i) Ground truth (iter=12).
(j) $e = 4.94\%$ (iter=12).
(k) Misclassification rate.
(a) Low-rank decomposition of class-based transformed training samples for subject3

(b) Low-rank decomposition of class-based transformed training samples for subject1

(c) class-based transformed testing samples for subject3

(d) class-based transformed testing samples for subject1

(a) Globally transformed testing samples for subject1

(b) Globally transformed testing samples for subject2
Random Forest

The ensemble model

Forest output probability $p(c|v) = \frac{1}{T} \sum_{t=1}^{T} p_t(c|v)$
Basic Formulation

- Node decision rule

\[
\arg \min_T ||TY_+||_* + ||TY_-||_* - ||T[Y_+, Y_-]||_*,
\]

\[s.t. ||T||_2 = 1,\]
Collapse and Separate

Original subspaces

Transformed subspaces

(a) $[\theta_{AB} = 0.79, \ \theta_{AC} = 0.79, \ \theta_{BC} = 1.05]$

$\mathbf{Y}_+ = \{A(green), B(blue)\}, \ \mathbf{Y}_- = \{C(red)\}.$

(b) $\mathbf{T} = \begin{bmatrix} 0.42 & 0.33 & -0.13 \\ 0.39 & 0.32 & -0.16 \\ -0.17 & -0.14 & 0.81 \end{bmatrix}$;

$[\theta_{AB} = 0.006, \ \theta_{AC} = 1.53, \ \theta_{BC} = 1.53].$

Original subspaces

Transformed subspaces

(c) $[\theta_{AB} = 1.05, \ \theta_{AC} = 1.05, \ \theta_{AD} = 1.05, \ \theta_{BC} = 1.32, \ \theta_{BD} = 1.39, \ \theta_{CD} = 0.53],$

$\mathbf{Y}_+ = \{A(blue), B(light blue)\}, \ \mathbf{Y}_- = \{C(yellow), D(red)\}.$

(d) $\mathbf{T} = \begin{bmatrix} 0.48 & 0.08 & -0.03 \\ 0.18 & 0.04 & -0.16 \\ -0.03 & -0.01 & 0.98 \end{bmatrix}$;

$[\theta_{AB} = 0.03, \ \theta_{AC} = 1.41, \ \theta_{AD} = 1.40, \ \theta_{BC} = 1.41, \ \theta_{BD} = 1.41, \ \theta_{CD} = 0.01].$
Results
Conclusion

• Learned Dictionaries

• Data as Features

• Learning Transforms

• Connections with Deep Learning
  – Theoretical Basis?
3.0: Thank You and Discussion

“If it works, there has to be math behind it”

I.D.

“If it works, there has to be math or biology behind it”

All work presented here has been done in collaboration with my students and collaborators: Sprechmann, Bronstein x2, Elhamifar, Qiang Qiu, Masci

Papers in arxiv.org