Task-Driven Sparse Coding for Classification of Motion Data

Babak Hosseini, Barbara Hammer

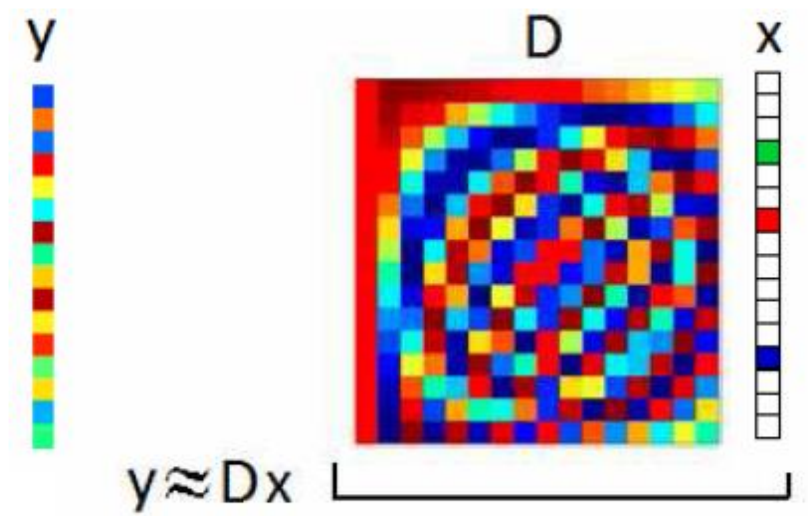
24. July. 2017

Bielefeld University
Cognitive Interaction Technology - Centre of Excellence

Sparse coding

- **Sparse coding:**
  - Providing a sparse representation of data
  - **Y**: Matrix of data vectors (columns)
  - **D:** a matrix of basic primitives (Dictionary)
  - **X**: Matrix of the Coefficient Vector
  - Sparsity constraints/objectives
    - 1-norm, cardinality, etc

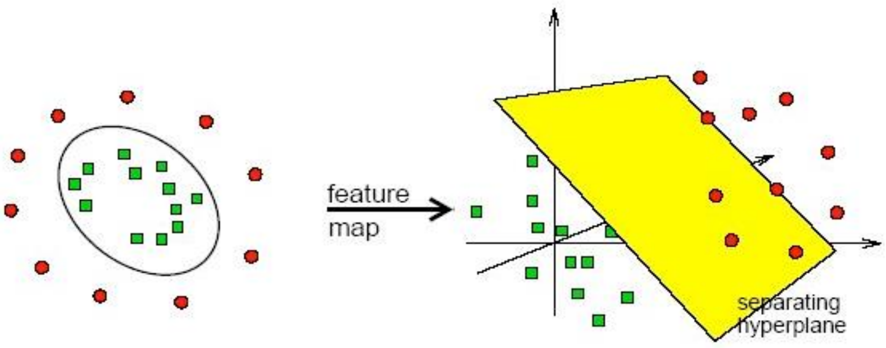
$$\min_{X,D} \quad \|\mathbf{Y} - \mathbf{DX}\|_F^2$$
$$\text{s.t} \quad \|\vec{X}_i\|_0 \leq T, \forall i = 1 \ldots N.$$



Classification tasks

- **Sparse coding → Classification:**
  - Classifier on top of sparse representation (***X***)
    - SVM
    - Linear/nonlinear classifier
    - nonlinear feature mapping

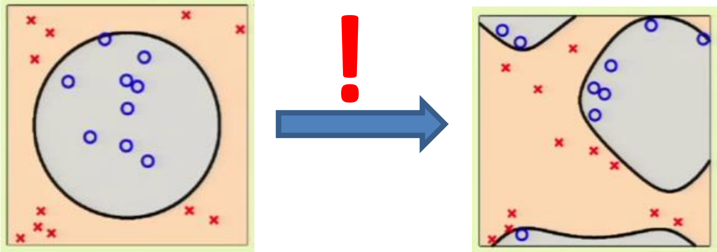
$$\min_{X,D} \quad \|\mathbf{Y} - \mathbf{D}\boxed{\mathbf{X}}\|_F^2$$
$$\text{s.t} \quad \|\vec{X}_i\|_0 \leq T, \forall i = 1...N.$$



Classification tasks

- **Sparse coding → Classification:**
  - Classifier on top of sparse representation (***X***)
    - SVM
    - Linear/nonlinear classifier
    - nonlinear feature mapping

$$\min_{X,D} \quad \|\mathbf{Y} - \mathbf{D}\boxed{\mathbf{X}}\|_F^2$$
$$\text{s.t} \quad \|\vec{X}_i\|_0 \leq T, \forall i = 1 \ldots N.$$



Classification tasks

- **Sparse coding → Classification:**
  - Classifier on top of sparse representation (**X**)
    - SVM
    - Linear/nonlinear classifier

$$\min_{X,D} \quad \|\mathbf{Y} - \mathbf{D}\boxed{\mathbf{X}}\|_F^2$$
$$\text{s.t} \quad \|\vec{X}_i\|_0 \leq T, \forall i = 1...N.$$

  - Augmenting the classifier to the main optimization:
  - ***f***: Classifier function/objective

$$\min_{\mathbf{X},\mathbf{W},\mathbf{D}} \|\mathbf{Y} - \mathbf{DX}\|_F^2 + \boxed{\lambda f(\mathbf{X},\mathbf{W})}$$
$$\text{s.t} \quad \|\mathbf{x}_i\|_0 \leq T$$

Universität Bielefeld

CITEC Cognitive Interaction Technology Center of Excellence Bielefeld University

Classification tasks

- **Task Driven dictionary learning**

$$\min_{X,D} f(X,D) \Rightarrow \text{Sparse Coding}$$

$$\min_{W} g(X,W) \Rightarrow \text{Classifier}$$

 - Bi-level optimization
 - Coupled

Classification tasks

- **Task Driven dictionary learning**

$$\min_{X,D} f(X,D) \qquad \Rightarrow \qquad \min_{X} f(X,D)$$

$$\min_{W} g(X,W) \qquad\qquad \min_{W,D} g(\boxed{X^*(D)},W)$$

  - If $X^*$is closed-form solution based on **D**
  - Augmenting {X,D} relationship into the 2nd optimization

Classification tasks

- **Task Driven dictionary learning**

$$\min_{X,D} f(X,D) \qquad \min_{W} g(X,W)$$

$$\Rightarrow \quad \min_{X} f(X,D) \qquad \min_{W,D} g(\boxed{X^*(D)},W)$$

$$\Rightarrow \quad X^* = \arg\min_{X} f(X,D) \qquad \{W^*,D^*\} = \arg\min_{W,D} g(D,W)$$

  - If X^* is closed-form solution based on **D**
  - Augmenting {X,D} relationship into the 2nd optimization
  - Optimizing D based on the classification task

Task driven framework

- **Task-driven sparse coding:**
- Each x_i^* use a different D_{I_i}
  - I_i: selected columns of D to reconstruct y_i
- Each x_i^* would result in a different $g_i(D_{I_i}, W)$
- $g(D, W) = \sum g_i(D_{I_i}, W)$
- !! Not a single structure
- Solution: Stochastic GD methods

$$X^* = \arg\min_X f(X, D)$$

$$\{W^*, D^*\} = \arg\min_{W,D} g(D, W)$$

$$\nabla_D g_i = \frac{\partial g_i}{\partial x_i} \frac{\partial x_i}{\partial D}$$

$$\nabla_D g \approx \sum_i \nabla_D g_i$$

$\Rightarrow$ Batch Optimization

Task driven framework

- **Alternating Optimization:**

- Solving wrt. *X* , D*, W* in a sequence in a loop*

$$X^* = \arg\min_X f(X, D)$$

$$\{W^*, D^*\} = \arg\min_{W,D} g(D, W)$$

Our task-driven framework

- **Task Driven Kernel Sparse Coding:**
  - Kernel space
    - $\Phi^*(\mathrm{D}) = \Phi(\mathrm{Y}) * A$
    - $A \in \mathbb{R}^{N \times k}$ in the input space
  - None-Negative framework
    - $\{X,A\}$ are positive ➔ interpretability
  - Linear classifier
    - H: training labels

$$\min_{\mathbf{X}} \ \|\varPhi(\mathbf{Y}) - \varPhi(\mathbf{Y})\mathbf{A}\mathbf{X}\|_F^2 \quad \text{s.t} \quad \mathbf{x}_{ij} \geq 0, \quad \|\mathbf{x}_i\|_0 \leq T$$

$$\min_{\mathbf{W},\boxed{\mathbf{A}}} \ \|\mathbf{H} - \mathbf{W}\boxed{\mathbf{X}}\|_F^2 + \lambda\|\mathbf{W}\|_F^2 \quad \text{s.t} \quad \mathbf{a}_{ij} \geq 0, \quad \|\mathbf{a}_i\|_2 = 1$$

Task Driven K-Sparse Coding

$$\min_{\mathbf{X}_i} \|\Phi(\mathbf{Y}_i) - \Phi(\mathbf{Y})\mathbf{A}\boxed{\mathbf{X}_i}\|_2^2$$

$$\text{s.t} \quad \mathbf{X}_i \geq 0, \quad \|\mathbf{X}_i\|_0 \leq T$$

- **Derivations:**
  - **nnKOMP solution :**

    $$x_i^* = (A_I^\top \mathcal{K} A_I)^{-1}(\mathcal{K}_{y_i} A_I)$$
  - I: selected columns from A
  - **Calculating** $\nabla_A(\|h_i - \mathbf{W}x_i\|_2^2)$ **:**

$$= 2[(x_i\rho)^\top + x_i\rho - (x_i s)^\top - (x_i s)]l + 2[\mathcal{K}_{y_i} s - \mathcal{K}_{y_i}\rho]^\top$$

$$s = [(\mathbf{A}_I^\top \mathcal{K} \mathbf{A}_I)^{-1}\mathbf{W}\mathbf{W}^\top x_i]^\top$$

$$\rho = h_i^\top \mathbf{W}(\mathbf{A}_I^\top \mathcal{K} \mathbf{A}_I)^{-1}$$

Task Driven K-Sparse Coding

- **Task Driven Algorithm:**

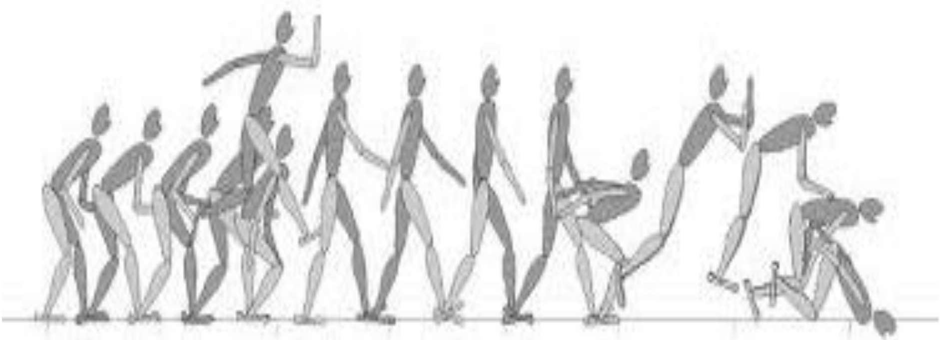
$$\min_{\mathbf{X}} \; \|\Phi(\mathbf{Y}) - \Phi(\mathbf{Y})\mathbf{A}\mathbf{X}\|_F^2$$
$$\text{s.t} \quad \mathbf{x}_{ij} \geq 0, \quad \|\mathbf{x}_i\|_0 \leq T$$

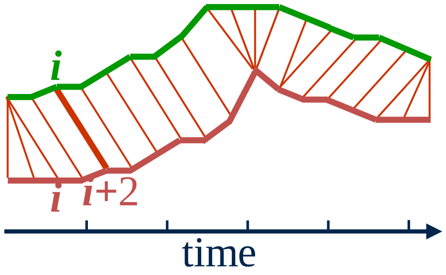
- **Loop till convergence:**
  - Finding ***X****: None-negative kernel OMP
  - Finding ***A****: stochastic projected gradient descent
  - Finding ***W****: linear programing (ridge regression)

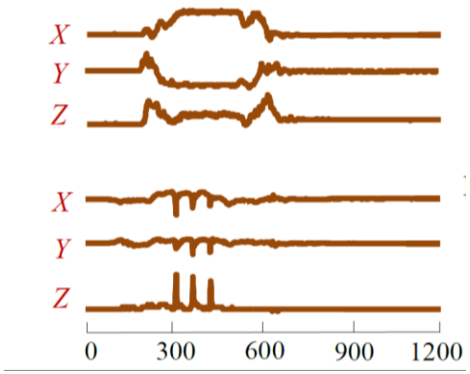
$$\min_{\mathbf{W},\mathbf{A}} \; \|\mathbf{H} - \mathbf{W}\mathbf{X}\|_F^2 + \lambda\|\mathbf{W}\|_F^2$$
$$\text{s.t} \quad \mathbf{a}_{ij} \geq 0, \quad \|\mathbf{a}_i\|_2 = 1$$

Application

- **Motion data classification:**
  - Multi-dimension Time-series
  - Kernel matrix:
    - Pair-wise similarity between the motions $\{y_i, y_j\}$
    - Using DTW distance







Universität Bielefeld

CITEC Cognitive Interaction Technology Center of Excellence Bielefeld University

Experiments

- **Results:**

	CMU		Cricket Signals		Articulatory Words		Squat	
	Acc	Rec. Err	Acc	Rec. Err	Acc	Rec. Err	Acc	Rec. Err
Task-KSC	93.17	5.72	86.6	10.2	98.88	10.5	100	1.02
LC-NNKSC	90.91	4.17	83.33	11.07	97.33	14.52	95	0.14
LC-KKSVD	86.36	7.44	83.33	10.1	97.33	7.8	85	3.4
NNKSC + SVM	82.12	2.26	80.74	7.3	95.28	5.3	85	1.8
K-Means+SVM	68	–	56.25	–	90	–	81	–
Affinity P.	90.1	–	68.75	–	92	–	100	–
K-PCA+SVM	50	–	56.25	–	60.66	–	37	–
kNN	86.36	–	79.16	–	96.66	–	100	–

Conclusion and Future works

- **Summary**
- Task-driven framework orients sparse coding towards the ***classification*** **objective**.
- The ***none-negative*** sparse representation improves classification performance.
- Non-negative kernel framework provides an ***interpretable*** model while classifying the data.

- **Future works:**
  - *Feature based classifier* via an additional parameter
  - *Online* version of the problem
  - Enhancing the *optimization strategy* → Speed, Robustness

Universität Bielefeld

CITEC
Cognitive Interaction Technology
Center of Excellence
Bielefeld University

- ***Thank you very much!***