Khanh-Hung TRAN (khanh-hung.tran@cea.fr)
Fred-Maurice NGOLE-MBOULA (LADIS)
Jean-Luc STARCK (DSM/IFRU)

Valencia - May 20-22 2018
ADA IX

DICTIONARY LEARNING WITH CLASSIFICATION

Why dictionary learning ?

Sparse dictionary learning (figure 1) has shown their super performance in data compression, noise reduction, … Further, it can deal with the classification by integrating with some other terms in objective function ($*$) like a linear classifier ($\mathbf{W}$) and graph regularized matrix (L_A) [1]. The latter term promotes the interactions between sparse codes ($\mathbf{A}$) to conserve the relations within data ($\mathbf{X}$). The feature dependencies is conserved thanks to the graph regularized matrix (L_D), so this problem called Dual Graph-Dictionary Learning (DG-DL). **Not like in traditional training, we use also test data in this stage** to regularize sparse codes, this helps to avoid over-learning because relations within all data (contains training and test) is conserved and reinforce the dictionary ($\mathbf{D}$) because we have more samples.

Index terms : Dictionary learning, Classification. Manifold learning.

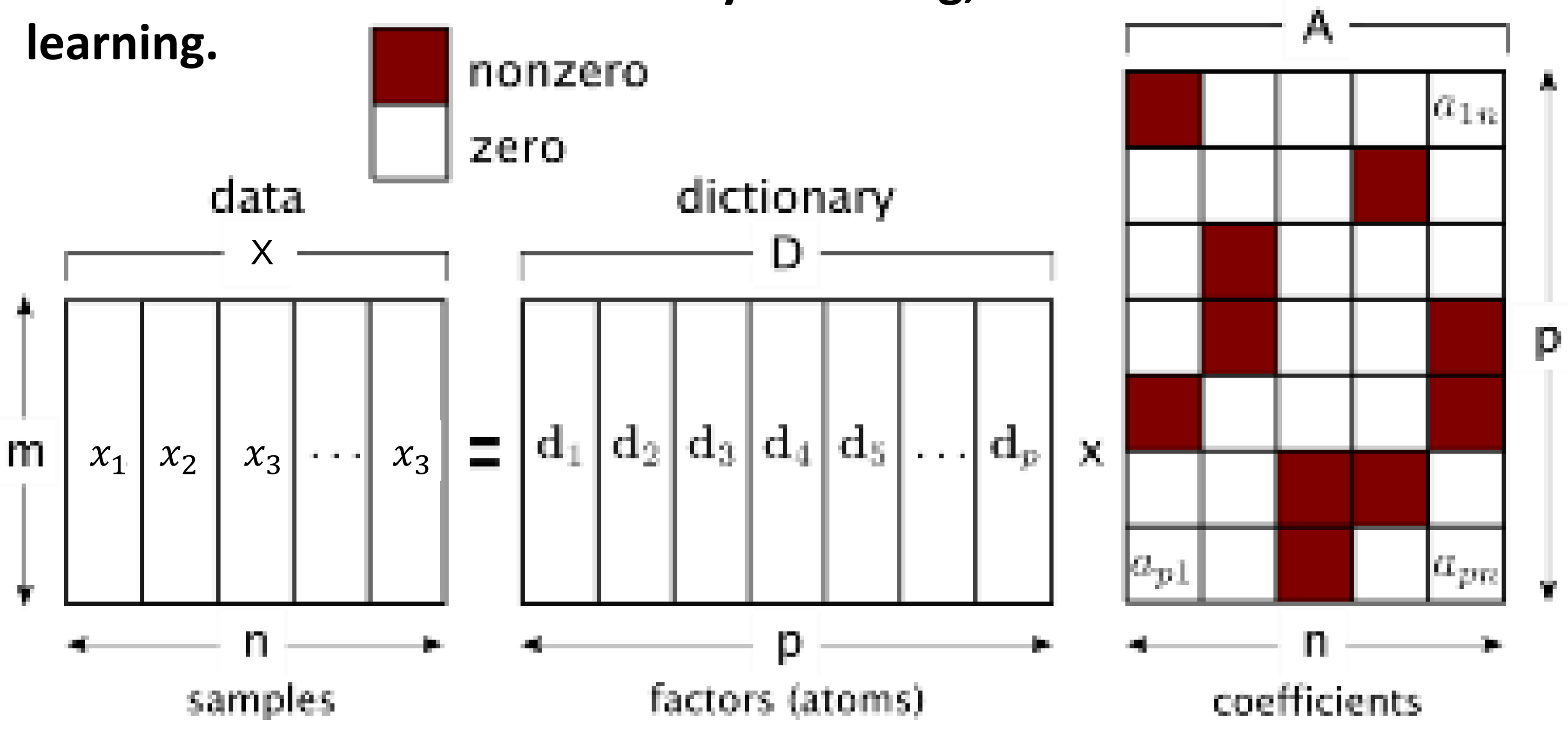

Figure 1. Sparse dictionary learning, un sample is reconstructed by a linear sum of atoms

$$\underset{\mathbf{W},\mathbf{A},\|\mathbf{D}[:,i]\|\le c}{\text{minimize}} \|\mathbf{X}-\mathbf{D}\mathbf{A}\|_F^2 + \alpha\,\mathrm{tr}\,(\mathbf{D}^\top L_D \mathbf{D}) + \beta\,\mathrm{tr}\,(\mathbf{A}L_A\mathbf{A}^\top) \qquad (*)$$
$$+\gamma\,\|\mathbf{Y}_{train}-\mathbf{W}\mathbf{A}_{train}\|_F^2 + \mu\,\|\mathbf{W}\|_F^2 + \lambda\,\|\mathbf{A}\|_1$$

Algorithm

Above objective function is solved iteratively by three stages (sparse coding, dictionary update and classifier update) with simple and efficient Proximal Splitting Method [2].

Repeat until convergence

Sparse coding :

$$\underset{\mathbf{A}}{\text{minimize}} \|\mathbf{X}-\mathbf{D}\mathbf{A}\|_F^2 + \beta\,\mathrm{tr}\,(\mathbf{A}L_A\mathbf{A}^\top)$$
$$+\gamma\,\|\mathbf{Y}_{train}-\mathbf{W}\mathbf{A}_{train}\|_F^2 + \lambda\,\|\mathbf{A}\|_1$$

Optimizing with Proximal Splitting Method with step descent $\left(2\left\|\mathbf{D}^\top\mathbf{D}\right\| + 2\beta\,\|L_A\| + 2\gamma\left\|\mathbf{W}^\top\mathbf{W}\right\|\right)^{-1}$

Dictionary update :

$$\underset{\|\mathbf{D}[:,i]\|\le c}{\text{minimize}} \|\mathbf{X}-\mathbf{D}\mathbf{A}\|_F^2 + \alpha\,\mathrm{tr}\,(\mathbf{D}^\top L_D\mathbf{D})$$

Optimizing with Proximal Splitting Method with step descent $\left(2\left\|\mathbf{A}\mathbf{A}^\top\right\| + 2\alpha\,\|L_D\|\right)^{-1}$

Classifier update :

$$\underset{W}{\text{minimize}} \|\mathbf{Y}_{train}-\mathbf{W}\mathbf{A}_{train}\|_F^2 + \mu/\gamma\,\|\mathbf{W}\|_F^2$$

Optimizing by finding zero of gradient

$$\mathbf{W} = \gamma\mathbf{Y}_{train}\mathbf{A}_{train}^\top\left(\gamma\mathbf{A}_{train}\mathbf{A}_{train}^\top + \mu I\right)^{-1}$$

return sparse code $\mathbf{A}$, classifier $\mathbf{W}$, dictionary $\mathbf{D}$

My works

We tested with database MINST and the precision is **88,54%** by using trained classifier ($\mathbf{W}$), and reach to **96.62%** if using SVM to reinforce with sparse code. We compare (DG-DL) with (DK-SVD), which just integrate the linear classifier ($\mathbf{W}$) in dictionary learning (figure 2). This result is very encouraging because it can be compared to neural network performance (about **99%**).

Our next job is to apply to astronomy data in which each image is 4Kx4K instead of 7x7 in MNIST and to carry on working with challenges below.

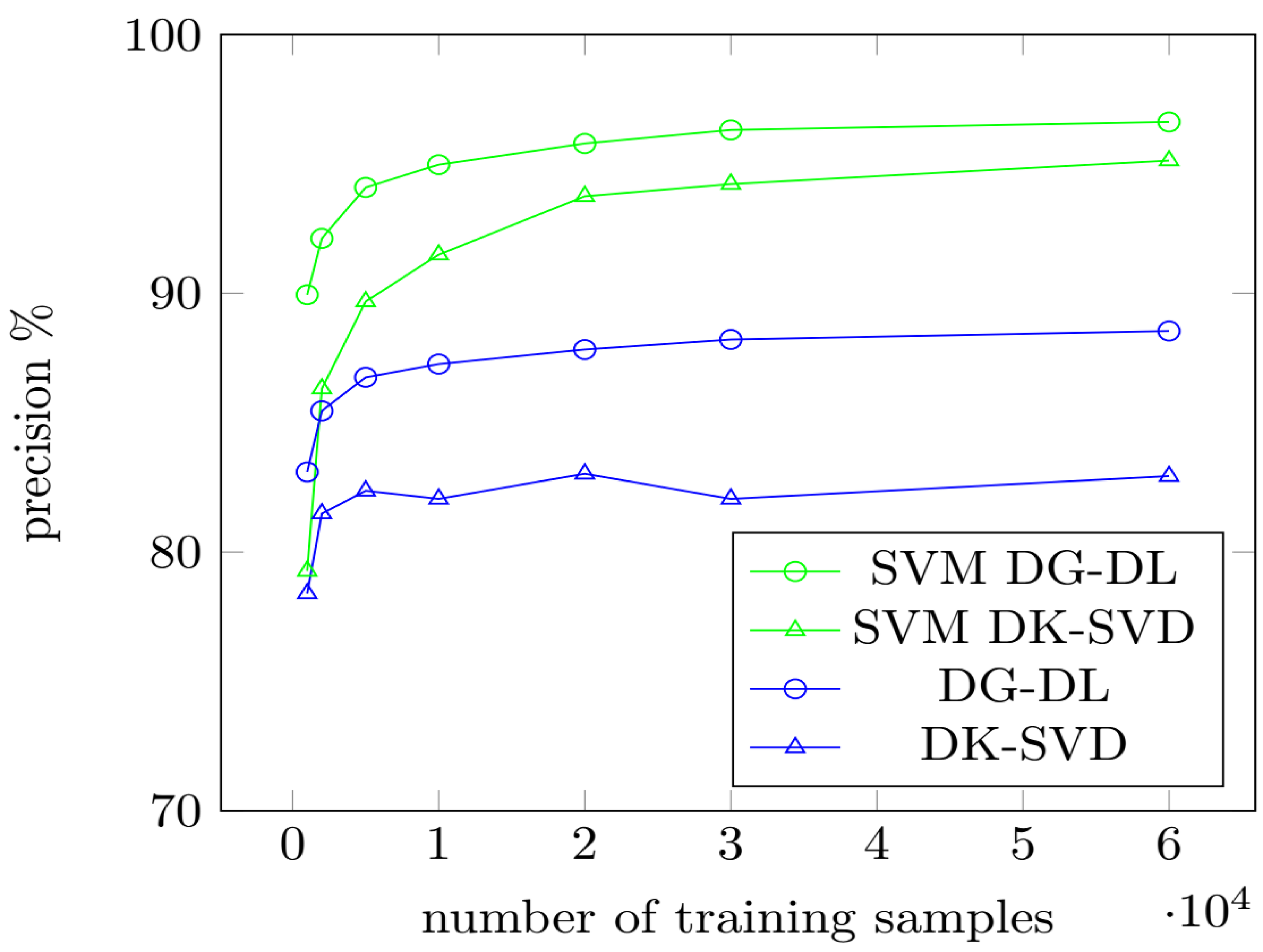

Figure 2. Comparison between DK-SVD and DG-DL, and boosted version with SVM

Challenge

Discriminant Power, which means to learn a dictionary to generate sparse codes which are more effectives in classification tasks but always in good condition of reconstruction.

Geometric Invariances, which means dictionary is able to learn the similar structures with different offsets (translations), orientations and scales in the training set.

Turning Problem which means to reduce manual turning tasks in parameters optimization of dictionary learning.

References

[1]. Y. Yankelevsky and M. Elad. Structure-aware classification using supervised dictionary learning. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 4421-4425, March 2017.

[2]. P. L. Combettes and J.-C. Pesquet. Proximal Splitting Methods in Signal Processing. ArXiv e-prints, December 2009.

[3]. Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y. Ng. Ecient sparse coding algorithms. In B. Sch?olkopf, J. Platt, and T. Homan, editors, Advances in Neural Information Processing Systems 19. MIT Press, Cambridge, MA, 2007.

[4]. Sam T. Roweis and Lawrence K. Saul. Nonlinear dimensionality reduction by locally linear embedding. SCIENCE, 290:2323{2326, 2000.