

3.2 Non-negative projections

Zhirong Yang, Zhijian Yuan, and Erkki Oja

Projecting high-dimensional input data into a lower-dimensional subspace is a fundamental research topic in signal processing, machine learning and pattern recognition. Non-negative projections are desirable in many real-world applications where the original data are non-negative, consisting for example of digital images or various spectra. It was pointed out by Lee and Seung [3] that the positivity or non-negativity of a linear expansion is a very powerful constraint, that seems to lead to sparse representations for the data. Their method, *non-negative matrix factorization (NMF)*, minimizes the difference between the data matrix \mathbf{X} and its non-negative decomposition \mathbf{WH} . The difference can be measured by the Frobenius matrix norm or the Kullback-Leibler divergence.

Yuan and Oja [7] proposed the *projective non-negative matrix factorization (PNMF)* method which replaces \mathbf{H} in NMF with $\mathbf{W}^T \mathbf{X}$, thus the data matrix \mathbf{X} is approximated as

$$\mathbf{X} \approx \mathbf{W}\mathbf{W}^T \mathbf{X}.$$

The nonnegative matrix \mathbf{W} is assumed to have a much lower rank than the data matrix itself. This actually combines the objective of principal component analysis (PCA) with the non-negativity constraint. The PNMf algorithm has been applied e.g. to facial image processing, and the empirical results indicate that PNMf is able to produce more spatially localized, part-based representations of visual patterns.

Recently, we have extended and completed the preliminary work with the following new contributions [5]: (1) formal convergence analysis of the original PNMf algorithms, (2) PNMf with the orthonormality constraint, (3) nonlinear extension of PNMf, (4) comparison of PNMf with two classical and two recent algorithms [6, 2] for clustering, (5) a new application of PNMf for recovering the projection matrix in a nonnegative mixture model, (6) comparison of PNMf with the approach of discretizing eigenvectors, and (7) theoretical justification of moving a term in the generic multiplicative update rule. Our in-depth analysis shows that the PNMf replacement has positive consequences in sparseness of the approximation, orthogonality of the factorizing matrix, decreased computational complexity in learning, close equivalence to clustering, generalization of the approximation to new data without heavy re-computations, and easy extension to a nonlinear kernel method with wide applications for optimization problems. Figure 3.1 demonstrates the advantage of PNMf over two other methods for the Nonnegative Kernel Principal Component Analysis problem.

Furthermore, we have proposed a more general method called α -PNMF [4], using α -divergence instead of Kullback-Leibler divergence as the error measure in PNMf. We have derived the multiplicative update rules for the new learning objective. The convergence of the iterative updates is proven using the Lagrangian approach. Experiments have been conducted, in which the new algorithm outperforms α -NMF [1] for extracting sparse and localized part-based representations of facial images.

Our method can also achieve better clustering results than α -NMF and ordinary PNMf for a variety of datasets. Table 3.1 shows the resulting clustering purities on six datasets.

References

- [1] Andrzej Cichocki, Hyekeyoung Lee, Yong-Deok Kim, and Seugjin Choi. Non-negative matrix factorization with α -divergence. *Pattern Recognition Letters*, 29:1433–1440,

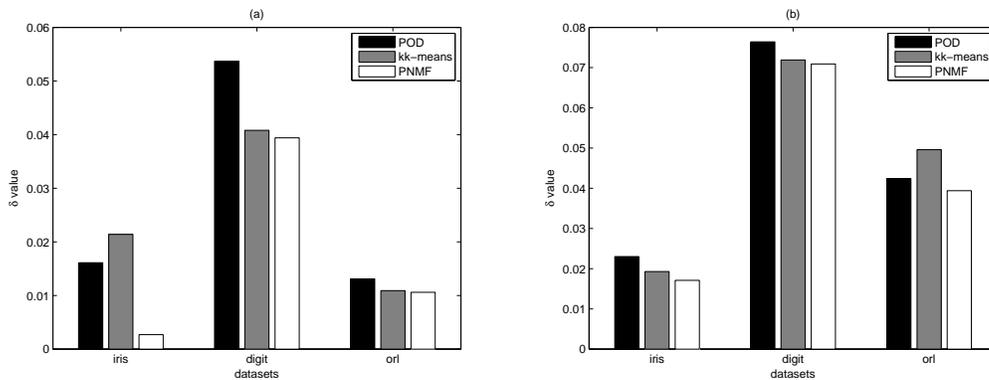


Figure 3.1: Comparison of POD, KK-means, and PNMF with (a) linear and (b) RBF kernels for the Nonnegative Kernel Principal Component Analysis problem. Smaller δ -values are better objectives relative to the KPCA solution

Table 3.1: Clustering purities using α -NMF, PNMF and α -PNMF. The best result for each dataset is highlighted with boldface font.

datasets	α -NMF			PNMF	α -PNMF		
	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$		-	$\alpha = 0.5$	$\alpha = 1$
Iris	0.83	0.85	0.84	0.95	0.95	0.95	0.97
Ecoli5	0.62	0.65	0.67	0.72	0.72	0.72	0.73
WDBC	0.70	0.70	0.72	0.87	0.86	0.87	0.88
Pima	0.65	0.65	0.65	0.65	0.67	0.65	0.67
AMLALL	0.95	0.92	0.92	0.95	0.97	0.95	0.92
ORL	0.47	0.47	0.47	0.75	0.76	0.75	0.80

2008.

- [2] Inderjit Dhillon, Yuqiang Guan, and Brian Kulis. Kernel kmeans, spectral clustering and normalized cuts. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 551–556, Seattle, WA, USA, 2004.
- [3] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.
- [4] Zhirong Yang and Erkki Oja. Projective nonnegative matrix factorization with α -divergence. In *Proceedings of 19th International Conference on Artificial Neural Networks (ICANN)*, pages 20–29, Limassol, Cyprus, 2009. Springer.
- [5] Zhirong Yang and Erkki Oja. Linear and nonlinear projective nonnegative matrix factorization. *IEEE Transaction on Neural Networks*, 2010. In press.
- [6] Stella X. Yu and Jianbo Shi. Multiclass spectral clustering. In *Proceedings of the Ninth IEEE International Conference on Computer Vision*, volume 2, pages 313–319, 2003.
- [7] Zhijian Yuan and Erkki Oja. Projective nonnegative matrix factorization for image compression and feature extraction. In *Proc. of 14th Scandinavian Conference on Image Analysis (SCIA 2005)*, pages 333–342, Joensuu, Finland, June 2005.