Comments

Comments on "Globally Maximizing, Locally Minimizing: Unsupervised Discriminant Projection with Application to Face and Palm Biometrics"

Weihong Deng, Jiani Hu, Jun Guo,
Honggang Zhang, and
Chuang Zhang

Abstract—In [1], UDP is proposed to address the limitation of LPP for the clustering and classification tasks. In this communication, we show that the basic ideas of UDP and LPP are identical. In particular, UDP is just a simplified version of LPP on the assumption that the local density is uniform.

Index Terms—Dimensionality reduction, unsupervised discriminant projection (UDP), locality preserving projection (LPP).

1 Equivalence of UDP and LPP

UDP [1] and LPP [2], [3], [4] are two typical linear extensions of the graph-motivated manifold learning methods. Let the matrix $X = [\mathbf{x}_1, \mathbf{x}_1, \cdots, \mathbf{x}_n]$ denote a set of n data points. The geometric structure of the data can be modeled by a weighted undirected graph $G = \{X, W\}$ with a vertex set X and an affinity matrix $W \in \mathbb{R}^{n\times n}$. The properties of the graph are characterized by the Laplacian matrix $L = D - W$, where D is a diagonal matrix $D_{ii} = \sum_j W_{ij}$. LPP is derived by the direct linear approximation of the Laplacian eigenmap [2], with its optimization as

$$\mathbf{w}_{lpp} = \arg\min_{\mathbf{w}} \frac{\mathbf{w}^T XLX^T \mathbf{w}}{\mathbf{w}^T XDX^T \mathbf{w}}. \tag{1}$$

In contrast, UDP is formulated by the geometric intuition of "globally maximizing, locally minimizing" [1], which defines an optimization problem as

$$\mathbf{w}_{udp} = \arg\min_{\mathbf{w}} \frac{\mathbf{w}^T S_L \mathbf{w}}{\mathbf{w}^T S_T \mathbf{w}}, \tag{2}$$

where S_L and S_T are the local and global scatter matrix, respectively [1].

From the criteria in (1) and (2), we realize that UDP and LPP essentially share the same basic idea, which is to emphasize the natural clusters in the projected data by simultaneously minimizing the local quantity and maximizing the global quantity. This comment is based on two simple observations. First, their local quantities, the numerators in (1) and (2), are two equivalent local scatters because we have

$$\mathbf{w}^T S_L \mathbf{w} = \frac{1}{n^2}\mathbf{w}^T XLX^T \mathbf{w}, \tag{3}$$

as shown in equation (13) of [1]. Note that this local scatter is a natural measure on the closeness of the local structure, which has been discussed in detail in [1] and [4]. Second, their global quantities, the denominators in (1) and (2), are two similar scatters that measure the separability of the whole data set. Specifically, UDP directly utilizes the global scatter of the classical discriminant analysis as its global quantity, which is equal to the mean square distance from any points to the global centroid, i.e.,

$$\mathbf{w}^T S_T \mathbf{w} = \frac{1}{n}\sum_i \left[\mathbf{w}^T(\mathbf{x}_i - \mathbf{m})\right]^2, \tag{4}$$

where $\mathbf{m} = \frac{1}{n}\sum_i \mathbf{x}_i$. Similarly, LPP characterizes its global quantity as the weighted sum of square distances from any points to the origin, i.e.,

$$\mathbf{w}^T XDX^T \mathbf{w} = \sum_i D_{ii}(\mathbf{w}^T \mathbf{x}_i)^2. \tag{5}$$

According to [5], the weighted mean of sample set, denoted as

$$\tilde{\mathbf{m}} = \frac{1}{(\sum_i D_{ii})}\left(\sum_i D_{ii}\mathbf{x}_i\right), \tag{6}$$

should be removed before LPP is applied. Thus, the global quantity of LPP can be interpreted as a weighted global scatter, i.e.,

$$\mathbf{w}^T \tilde{X} D \tilde{X}^T \mathbf{w} = \sum_i D_{ii}\left[\mathbf{w}^T(\mathbf{x}_i - \tilde{\mathbf{m}})\right]^2, \tag{7}$$

where $\tilde{X}$ is the centered version of X, i.e.,

$$\tilde{X} = [\mathbf{x}_1 - \tilde{\mathbf{m}}, \cdots, \mathbf{x}_n - \tilde{\mathbf{m}}].$$

It would be important to note that the only difference between LPP and UDP comes from the degrees (D_{ii}) of the vertices, which physically represent the local density around the data points. If the local density around each point is equal, i.e., $\forall i,\ D_{ii} = \rho$, ρ is a constant, we have $\mathbf{m} = \tilde{\mathbf{m}}$ and

$$\mathbf{w}^T S_T \mathbf{w} = \frac{1}{n\rho}\mathbf{w}^T \tilde{X} D \tilde{X}^T \mathbf{w}. \tag{8}$$

Based on (3) and (8), it can be concluded that the projections derived from UDP and LPP are identical under the assumption that the local density is uniform.

Certainly, as this assumption is rarely held in the real-world applications, there is always a slight difference[1] between their performance due to the nonuniformity of D_{ii}. However, one can clearly see from this section that LPP and UDP are equivalent in term of basic ideas and geometric intuitions.

2 Statement of the Problem

In [1], the authors overlook the globally maximizing character of LPP so that they wrongly understand that LPP would preserve the local structure by simply minimizing the local scatter, without any concern on the global scatter. They claim that "the consideration of the nonlocal (global) quantity makes UDP more intuitive and more powerful than LPP for classification or clustering tasks." However, as shown in Section 1, LPP considers

1. In practice, the commonly used k-nearest neighbor graph tends to be regular and most vertices have approximately the same degree. Thus, UDP and LPP are very similar to each other, and work almost equally.

• The authors are with the Laboratory of Pattern Recognition and Intelligent System, School of Information Engineering, Beijing University of Posts and Telecommunications, PO Box 186, 100876, Beijing, China.
E-mail: {whdeng, guojun, zhhg}@bupt.edu.cn, cughu@126.com, zcbupt@sina.com.

Manuscript received 14 May 2007; revised 19 Aug. 2007; accepted 27 Aug. 2007; published online 11 Sept. 2007.
Recommended for acceptance by L. O'Gorman.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2007-05-0282.
Digital Object Identifier no. 10.1109/TPAMI.2007.70783.

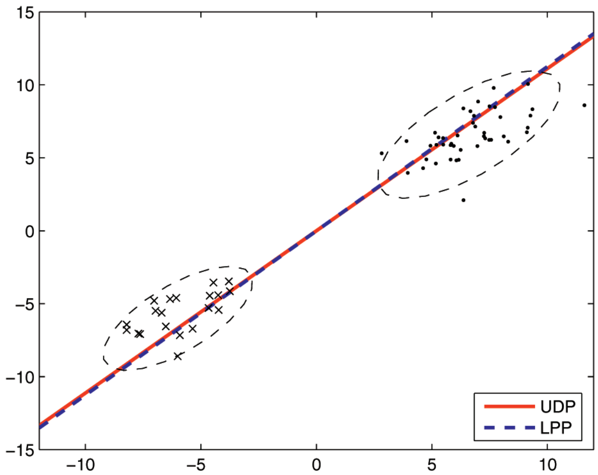

Fig. 1. Two clusters of samples in two-dimensional space and the projection directions of UDP and LPP.

the global scatter in a manner which is more generalized than UDP. In particular, UDP is just a simplified version of LPP on the assumption of uniform local density.

One should be aware that Fig. 1 of [1] and the accompanying explanation on p. 651 have given a misleading comparison between UDP and LPP. The authors illustrate that UDP successfully discriminates the two clusters, whereas LPP collapses all samples from two clusters together. However, by randomly generating two Gaussian (not uniformly distributed) clusters and running the two algorithms to simulate their descriptions, we find that LPP always derives the discriminative direction similar to that from UDP, as shown in Fig. 1. One can see from the figure that there is no meaningful difference between UDP and LPP.

The authors of [1] also ignore the physical meaning of D_{ii} (local density) and argue that "maximizing $\mathbf{w}^T S_D \mathbf{w}$ ($S_D = \tilde{X} D \tilde{X}^T$) does not make sense with respect to discrimination" on pp. 656. Note that D_{ii} is a natural measure on the "importance" of the data points [3] since the points with large D_{ii} tend to be the representative (central) points of the clusters and the ones with small D_{ii} are likely the noise and outliers. Hence, maximizing the D_{ii}-weighted global scatter of LPP, i.e., $\mathbf{w}^T S_D \mathbf{w}$, explicitly emphasizes the natural clusters with strong noise robustness and thus makes clear sense to discrimination. Furthermore, from the perspective of manifold learning, D_{ii} provides a discrete approximation to the standard measure on a Riemannian manifold and plays an indispensable role when LPP finds the optimal linear approximations to the eigenfunctions of the Laplace Beltrami operator on the manifold,[2] Conversely, the global scatter of UDP, i.e., $\mathbf{w}^T S_T \mathbf{w}$, ignores the effects of D_{ii}, weighting each point equally. Strictly speaking, this quantity makes UDP relatively sensitive to outliers and noise and imperfect with respect to the theory of manifold learning, though it is also intuitive for discrimination.

One may wonder why UDP can outperform LPP in the experiments of [1] in despite of its theoretical deficiency. Note that the experiments in [1] mainly focus on the special cases with small (or nonrepresentative) training data set, where the algorithms have to learn from undersampled distributions. In these experiments, the local density D_{ii} of LPP becomes unreliable, whereas UDP performs an useful regularization by the "uniform density" assumption and thus generalizes better when compared against LPP. Additional empirical comparisons of UDP and LPP are necessary for more comprehensive understanding, but this is beyond the scope of this comment paper.

2. The authors would like to acknowledge the anonymous reviewer for suggesting this point.

Finally, we would like to conclude that UDP is an effective algorithm as a simplified, or regularized, version of LPP, but there is no reason to prefer UDP over LPP for the general classification and clustering tasks.

ACKNOWLEDGMENTS

The authors would like to thank the associate editor and the anonymous reviewers for their careful reading and constructive suggestions. This work was partially sponsored by the NSFC (Natural Science Foundation of China) under Grant No.60675001 and the Foundation of China Education Ministry for Century Spanning Talent.

REFERENCES

[1] J. Yang, D. Zhang, J.-Y. Yang, and B. Niu, "Globally Maximizing, Locally Minimizing: Unsupervised Discriminant Projection with Applications to Face and Palm Biometrics," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 29, no. 4, pp. 650-664, Apr. 2007.
[2] X. He and P. Niyogi, "Locality Preserving Projections," *Proc. Conf. Advances in Neural Information Processing System*, 2003.
[3] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face Recognition Using Laplacianfaces," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, no. 3, pp. 328-340, Mar. 2005.
[4] X. He, D. Cai, and W. Min, "Statistical and Computational Analysis of Locality Preserving Projection," *Proc. 22nd Int'l Conf. Machine Learning,* pp. 281-288, 2005.
[5] D. Cai, X. He, and J. Han, "Document Clustering Using Locality Preserving Indexing," *IEEE Trans. Knowledge and Data Eng.,* vol. 17, no. 12, pp. 1624-1637, Dec. 2005.