The Connection Between Manifold Learning and Distance Metric Learning

Liu Yang

October 28, 2007

Manifold Learning learns a low-dimensional embedding of the latent manifold. In this report, we give the definition of distance metric learning, provide the categorization of manifold learning, and describe the essential connection between manifold learning and distance metric learning, with special emphasis on nonlinear manifold learning, including ISOMAP, Laplacian Eigenamp (LE), and Locally Linear Embedding (LLE).

1 Distance Metric Learning and Linear Projective Mapping

The distance between points $\mathbf{x} \in R^M$ and $\mathbf{y} \in R^M$ is defined as

$$d(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^\top \mathbf{A} (\mathbf{x} - \mathbf{y}) \tag{1}$$

The typical problem of distance metric learning is the learning of the $\mathbf{A} \in \Re^{M \times M}$.

We can further write Equation (1) as

$$\begin{aligned} d(\mathbf{x}, \mathbf{y}) &= (\mathbf{x} - \mathbf{y})^\top \mathbf{A}^{\frac{1}{2}\top} \mathbf{A}^{\frac{1}{2}} (\mathbf{x} - \mathbf{y}) \\ &= (\mathbf{A}^{\frac{1}{2}}\mathbf{x} - \mathbf{A}^{\frac{1}{2}}\mathbf{y})^\top (\mathbf{A}^{\frac{1}{2}}\mathbf{x} - \mathbf{A}^{\frac{1}{2}}\mathbf{y}) \\ &= (\mathbf{P}\mathbf{x} - \mathbf{P}\mathbf{y})^\top (\mathbf{P}\mathbf{x} - \mathbf{P}\mathbf{y}) \end{aligned} \tag{2}$$

with $\mathbf{P} = \mathbf{A}^{\frac{1}{2}}$. It is clear that the learning of $\mathbf{A}$ is equivalent to the learning of a linear projective mapping $\mathbf{P}$ in the feature space.

The linear manifold learning methods, which learn a linear transformation, can be interpreted as learning the projective $\mathbf{A}^{\frac{1}{2}}$ matrix as above, and essentially solving the exact problem as distance metric learning. Therefore, any linear manifold learning algorithm that is able to learn an explicit projective mapping has the equivalent goal of learning a distance metric.

Methods	Linear	Nonlinear
Global structure preserved	Principal Component Analysis (PCA) [4] Multidimensional Scaling (MDS) [3] Independent Components Analysis (ICA) [2]	ISOMAP [8]
Local structure preserved	Locality Preserving Projections (LPP) [5] Neighborhood Preserving Embedding (NPE) [6]	Laplacian Eigenamp (LE) [1] Locally Linear Embedding (LLE) [7]

Table 1: Categorization of Manifold Learning Methods

2 Manifold Learning Methods and their connections to Distance Metric Learning

Manifold Learning approaches can be categorized along the following two dimensions: first, the learnt embedding is linear or nonlinear; and second, the structure to be preserved is global or local (see Table 1). Based on the analysis in section 1, all the linear methods in Table 1 except Multidimensional Scaling (MDS), learn an explicit linear projective mapping and can be interpreted as the problem of distance metric learning. MDS finds the low-rank projection that best preserves the inter-point distance matrix $\mathbf{E}$. This low-rank projection has its intrinsic relation to the PCA linear mapping : $V^{\text{pca}} = XV^{\text{mds}}$. Here X is the data matrix, V^{pca} are the principle eigenvectors of the PCA covariance matrix; and V^{mds} are the principle eigenvectors of $\mathbf{HEH}$ ($\mathbf{H}$ is the centering matrix).

The nonlinear manifold learning methods, with no explicit projective mapping to be learnt, generate nonlinear Embedding (e.g. ISOMAP, LLE, and LE). To analysis their connection to distance metric learning, first, we should realize the common nature of preserving distance constraints, although the specific forms of distance constraints may vary. Distance metric learning methods preserve the binary distance (1 as must-link and 0 as cannot-link), for instance, pairwise constraints and chunklets. ISOMAP preserves the geodesic distance between pairs of data points, which is estimated by computing shortest paths through large sublattices of data (ISOMAP applies MDS to the geodesic distance matrix). LLE preserves distance based on locally linear combination of neighborhood. And LE preserves the distance described by a weighted connected graph constructed from neighborhood. Second, the nonlinear representation computed by ISOMAP, LLE, and LE, can be interpreted as the data representation $(\mathbf{x}_i - \mathbf{x}_j)\mathbf{A}^{\frac{1}{2}}$ obtained in distance metric learning. Third, and more importantly, linear mappings that approximate Locally Linear Embedding (LLE) and Laplacian Eigenamp can be computed. In particular, Neighborhood Preserving Embedding (NPE) is a linear approximation of Locally Linear Embedding (LLE); and Locality Preserving Projection (LPP) is a linear approximation to Laplacian Eigenmaps (LE). Below provides the details.

Denote $\mathbf{W}$ as the weight matrix, $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_N) \in \Re^{M\times N}$ as the data matrix containing N data points in the original feature space, and $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_N) \in \Re^{m\times N}$ as the nonlinear embedding matrix, with $m \leq M$.

Locally Linear Embedding (LLE) constructs a neighbor-preserving mapping by minimizing the cost function $\Phi(\mathbf{Y}) = \sum_i \|\mathbf{y}_i - \sum_{i=1}^{K} \mathbf{W}^*_{ij} y_{ij}\| = \|\mathbf{Y}^\top \mathbf{M} \mathbf{Y}\|^2$, where $\mathbf{M} = (\mathbf{I} - \mathbf{W}^*)^\top (\mathbf{I} - \mathbf{W}^*)$, and $\mathbf{W}^* = \text{argmin} \sum_i \|\mathbf{x}_i - \sum_j W_{ij} \mathbf{x}_j\|^2$. Neighborhood Preserving Embedding (NPE) introduces a linear transformation $\mathbf{B} \in \Re^{M \times m}$ so that $\mathbf{Y} = \mathbf{B}^\top \mathbf{X}$. Then the above minimization problem reduces to finding $\mathbf{B} = \text{argmin}_{\mathbf{B}^\top \mathbf{X} \mathbf{X}^\top \mathbf{B} = 1} \mathbf{B}^\top \mathbf{X} \mathbf{M} \mathbf{X}^\top \mathbf{B}$. The transformation matrix $\mathbf{B}$ is the minimum eigen solution to the generalized eigenvector problem: $\mathbf{X} \mathbf{M} \mathbf{X}^\top \mathbf{B} = \lambda \mathbf{X} \mathbf{X}^\top \mathbf{B}$. The matrix M provides a discrete approximation to the Laplace Beltrami operator on the manifold [6]. This indicates NPE provides a way to linearly approximate the eigenfunctions of the Laplace Beltrami operator on the manifold.

Laplacian Eigenamp (LE) computes the nonlinear embedding $\mathbf{y}$ by solving the general eigen problem $\mathbf{L}\mathbf{y} = \lambda \mathbf{D}\mathbf{y}$, where $\mathbf{D}$ is a diagonal matrix whose entries are column sums of the weight matrix $\mathbf{W}$, and $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the Laplacian matrix. Locality Preserving Projections (LPP) introduces a linear transformation $\mathbf{C}\Re^{M \times m}$, so that $\mathbf{y} = \mathbf{C}^\top \mathbf{x}$. Then the eigen problem in LE can be reduced to the solution of a generalized eigen problem $\mathbf{X}\mathbf{L}\mathbf{X}^\top \mathbf{C} = \lambda \mathbf{X}\mathbf{D}\mathbf{X}^\top \mathbf{C}$.

From the above analysis, we can see that LLE and LE are both associated with distance metric learning through their linear approximation.

Conclusively, linear manifold learning is solving the similar problem as distance metric learning; and nonlinear manifold learning also has its essentially connections to distance metric learning.

References

[1] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6), 2003.

[2] P. Common. Independent component analysis — a new concept? *Signal Processing*, 36:287–314, 1994.

[3] T. Cox and M. Cox. *Multidimensional Scaling*. Chapman and Hall, 1994.

[4] R. Gonzales and R. Woods. *Digital Image Processing*. Addison-Wesley, 1992.

[5] X. He and Partha Niyogi. Locality preserving projections. In *NIPS*, 2003.

[6] Xiaofei He, Deng Cai, Shuicheng Yan, and Hong-Jiang Zhang. Neighborhood preserving embedding. In *Proc. Int. Conf. on Computer Vision*, 2005.

[7] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. In *Science*, 2000.

[8] J.B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290, 2000.