Sparse Representation Based Projections

Radu Timofte[1]
http://homes.esat.kuleuven.be/~rtimofte
Luc Van Gool[1,2]
Luc.VanGool@esat.kuleuven.be

[1]ESAT-PSI-VISICS / IBBT
Katholieke Universiteit Leuven, Belgium
[2]BIWI
ETH Zurich, Switzerland

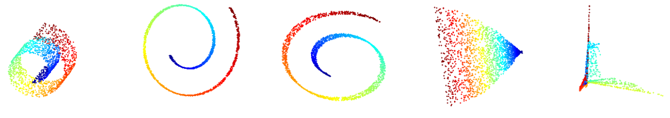
original LPP SRLP LLE SRE

Figure 1: SwissRoll 2D projections form 3D

We adapt the Locality Preserving Projections (LPP)[4] and Locally Linear Embedding (LLE)[5] techniques to preserve the sparse representation property in the embedded space. The resulting techniques are Sparse Representation based Linear Projections (SRLP) and Sparse Representation based Embedding (SRE). We compare them to the original methods on several benchmarks, dealing with faces, traffic signs, digits, for both unsupervised and supervised cases.

Most signals have a sparse representation as a linear combination of a reduced subset of signals from the same space naturally biased towards their own class. This is the starting point for Sparse Representation based Classification (SRC)[6]. The sparsity and compressed sensing idea brought new formulations such as Sparse PCA[7] or Sparse Regression Discriminant Analysis (SRDA)[1], which aim at having representations which are sparse over the basis directions in the embeddings.

In an image-based recognition task we have a set of roughly aligned labeled training images $\{\mathbf{x}_i, l_i\}$ from C classes. $\{\mathbf{x}_i \in \mathbb{R}^M\}$ is the vectorial representation (here M grayscale pixel values), while $l_i \in \{1 \dots C\}$ gives the class of the i-th image. We are searching a D-dimensional space such that the corresponding points $\{\mathbf{y}_i \in \mathbb{R}^D\}$ preserve the sparse representation property as defined next. Let $\mathbf{X}_{N\times M} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N]^T$, $\mathbf{Y}_{N\times D} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N]^T$, and N be the number of training samples.

For each point $\mathbf{x}_i$ we are searching for its **sparse representation**

$$\text{minimize} \quad \|\mathbf{w}_i\|_1 \quad \text{subject to} \quad \|\mathbf{x}_i - \textstyle\sum_{j=1, j\neq i}^{N} \mathbf{w}_{ij}\mathbf{x}_j\|_2 \leq \varepsilon \qquad (1)$$

where $\mathbf{w}_i \in \mathbb{R}^N$ is the sparse vector of weights, $\mathbf{w}_{ij}$ shows the contribution of the sample $\mathbf{x}_j$ to the sparse representation of $\mathbf{x}_i$ and $\varepsilon \in \mathbb{R}$ is the measurement noise. For Compressed Sensing, l_1-minimization proved to be efficient in recovering the sparsest solutions to underdetermined systems of linear equations. We are using the homotopy [2] method for solving (1). In a supervised sparse representation scenario we can compute (1) restricted to samples sharing the same label.

Sparse Representation-based Linear Projections (SRLP) replaces the graph representation from LPP [4] with a matrix of weights coming from the sparse representations $\mathbf{w}_i$ that are computed for each training sample $\mathbf{x}_i$. The SRLP weighted symmetric matrix is

$$\mathbf{W}_{N\times N} = \max\{[|\mathbf{w}_1|, |\mathbf{w}_2|, \cdots, |\mathbf{w}_N|], [|\mathbf{w}_1|, |\mathbf{w}_2|, \cdots, |\mathbf{w}_N|]^T\} \qquad (2)$$

The eigenmap step is common to the original LPP and SRLP and consists of computing the eigenvectors and eigenvalues for

$$\mathbf{XLX}^T\mathbf{a} = \lambda\mathbf{XDX}^T\mathbf{a} \qquad (3)$$

where $\mathbf{D}$ is a diagonal matrix whose entries are column sums of $\mathbf{W}$, $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ji}$. $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the Laplacian matrix. The eigenvectors of the smallest eigenvalues yield the projection matrix.

Sparse Representation-based Embedding (SRE) replaces for each training sample $\mathbf{x}_i$ the K-nearest neighbors of LLE by those yielding a sparse representation, and the reconstruction weights by the coefficients $\mathbf{w}_i$ from that representations . In LLE the row weights $\mathbf{w}_i^T$ are normalized by dividing by the sum of their elements, and so we do for SRE, $\mathbf{w}'_i = \mathbf{w}_i / (\sum_{j=1}^{N} \mathbf{w}_{ij})$. The SRE weighted adjacency matrix is

$$\mathbf{W}_{N\times N} = [\mathbf{w}'_1, \mathbf{w}'_2, \cdots, \mathbf{w}'_N] \qquad (4)$$

The embedding step is common between LLE and SRE and consists of minimizing the following cost function in the embedding space:

$$\Phi(\mathbf{Y}) = \sum_{i=1}^{N} \|\mathbf{y}'_i - \sum_{j=1, j\neq i}^{N} \mathbf{w}'_{ij}\mathbf{y}_j\|^2 \qquad (5)$$

Table 1: Accuracy[%], unsupervised case (best dim. in subscript)

	classifier	RLPP	RSRLP	LLE	SRE	PCA
PIE	NN	95.7_{190}	96.5_{101}	89.8_{200}	93.5_{190}	93.9_{198}
	${}_{l}$SVM	91.3_{99}	94.3_{97}	87.1_{190}	93.3_{200}	95.5_{194}
	${}_{ik}$SVM	95.5_{99}	96.4_{99}	89.9_{180}	94.6_{200}	97.6_{196}
	${}_{rbf}$SVM	95.1_{59}	97.1_{99}	90.3_{200}	94.4_{200}	97.9_{150}
GTSRB	NN	80.3_{70}	84.3_{74}	61.0_{200}	66.5_{170}	64.8_{180}
	SRC	85.5_{73}	88.8_{78}	62.6_{200}	68.9_{180}	72.6_{99}
	${}_{poly}$SVM	90.4_{58}	93.1_{78}	62.4_{200}	70.6_{180}	86.9_{115}
	${}_{rbf}$SVM	89.4_{73}	93.0_{78}	63.9_{200}	71.3_{180}	86.8_{100}

Table 2: Accuracy[%], supervised case (best dim. in subscript)

	classifier	RLPP	RSRLP	LLE	SRE	RLDA
PIE	NN	97.7_{55}	97.8_{37}	92.8_{150}	95.9_{170}	97.5_{51}
	${}_{l}$SVM	96.0_{99}	97.1_{95}	91.8_{180}	96.1_{180}	97.3_{61}
	${}_{ik}$SVM	97.4_{94}	97.6_{63}	92.6_{180}	94.5_{160}	97.3_{65}
	${}_{rbf}$SVM	97.6_{79}	98.0_{55}	92.8_{200}	96.0_{160}	97.6_{67}
GTSRB	NN	87.0_{48}	91.8_{28}	67.1_{170}	72.3_{50}	92.7_{40}
	SRC	90.0_{40}	93.6_{53}	67.5_{200}	71.1_{100}	93.5_{42}
	${}_{poly}$SVM	92.0_{34}	94.7_{78}	67.4_{200}	72.0_{80}	92.6_{32}
	${}_{rbf}$SVM	93.5_{79}	94.6_{78}	68.4_{200}	72.1_{80}	92.9_{42}

The solution is obtained by solving the eigenvector problem of a sparse matrix $\mathbf{M} = (\mathbf{I} - \mathbf{W})^T(\mathbf{I} - \mathbf{W})$, where $\mathbf{I}$ is the identity matrix of rank N. The embedding coordinates are found as the eigenvectors with the smallest non-zero eigenvalues.

In the **experiments** (partial results in Table 1 and 2) we use a battery of classifiers and a subset of the CMU PIE face dataset with 11554 images, the BTSC (7125 images) and GTSRB (39209) traffic sign datasets, the MNIST (70000) digit dataset, and the SwissRoll (2000) dataset. For all the image datasets, the images are cropped, resized to 28 × 28, and the feature vectors are the pixel values (256 gray levels) l_2-norm normalized.

Regularization improves the performance for the LDA [3] and LPP techniques for an appropriate weighting parameters, here set to 10^{-2}. The number of nearest neighbors for LLE is set to 12.

RSRLP improves over RLPP for the face and traffic sign datasets for all considered classifiers, while on MNIST digits it is on par with RLPP. SRE improves over LLE, but both are outperformed by the linear algorithms. Tuning the number of neighbors of LLE for each setting (dimensionality of the projection, dataset, and classifier) could improve performance, but is inefficient and cumbersome. SRE and SRLP, on the other hand, have no such tuning parameter.

The proposed methods – SRLP and SRE – are on par or consistently outperform the original formulations in supervised and unsupervised learning settings. The sparse representation property shows great potential.

[1] Deng Cai. *Spectral Regression: A Regression Framework for Efficient Regularized Subspace Learning*. PhD thesis, University of Illinois at Urbana-Champaign, May 2009.

[2] David L. Donoho and Yaakov Tsaig. Fast solution of l_1-norm minimization problems when the solution may be sparse. *IEEE Transactions on Information Theory*, 54(11):4789–4812, 2008.

[3] J.H. Friedman. Regularized discriminant analysis. *Journal of the American Statistical Association*, 84:165–175, 1989.

[4] Xiaofei He and Partha Niyogi. Locality preserving projections. In *NIPS*, 2003.

[5] Sam Roweis and Lawrence Saul. Nonlinear dimensionality reduction by locally linear embedding. In *ICCV*, 2001.

[6] J. Wright, A. Y. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *PAMI*, 31(2), 2009.

[7] Hui Zou, Trevor Hastie, and Rob Tibshirani. Sparse principal component analysis. *JCGS*, 15:262–286, 2006.