RESEARCH STATEMENT

Yingzhen Yang
E-mail: superyyzg@gmail.com
Personal Website: `yingzhenyang.com`

1 Research Overview

The era of big data arouses the needs of sophisticated machine learning methodologies for data analysis, and it is particularly interesting to develop provable statistical machine learning methods to analyze large-scale and complex high-dimensional data in various real scenarios in an efficient and effective manner. My research focuses on theory and application of learning low-dimensional structure in high-dimensional data, as well as provable optimization algorithms on large-scale data. From this angle, my research covers an extensive range of important areas of machine learning, including subspace learning, manifold learning, sparse representation, kernel methods, deep learning, and optimization for machine learning. Low-dimensional structure widely exists in high-dimensional data. For example, the visual data exhibit low-dimensional structures due to repetitive texture or structural patterns, redundant sampling and transformations of the same object, etc.

Machine learning methods which exploit the low-dimensional structure in high-dimensional data have been extensively studied in the literature. The representative low-dimensional linear structure in high-dimensional data is subspace, and the typical example of low-dimensional nonlinear structure is manifold. As fundamental problems of machine learning, subspace learning and manifold learning have received extensive attention in the machine learning literature. For example, the well-known Principal Component Analysis (PCA) works perfectly if the data are distributed around a single subspace. Moreover, it is a widely adopted assumption that high-dimensional scientific data lie on or close to subspaces or submanifolds of low intrinsic dimension in many scientific disciplines, and various learning tasks on such data benefit from subspace or manifold learning which identifies the underlying subspace or manifold structure in the data.

While subspace learning and manifold learning have been extensively studied in the machine learning literature, there is few research that studies provable learning methods under mild assumptions in this direction and the application of learning such low-dimensional structure in deep learning. To this end, we propose provable subspace learning method under mild assumptions and novel manifold learning method coupled with application in deep learning. Our provable subspace learning method, ℓ^0-Sparse Subspace Clustering (ℓ^0-SSC), employs a novel ℓ^0-induced sparse model to recover the underlying subspace structure under much milder assumptions compared to ℓ^1-SSC, its ℓ^1 counterpart and a representative sparse subspace clustering method. ℓ^0-SSC has been regarded as a well-known and important work in subspace learning literature. We also propose novel manifold learning methods to learn regularized sparse graph and regularized sparse coding in accordance with the manifold structure of the data. The regularized sparse coding method features a neural network as a fast encoder for approximating the sparse codes, and this encoder reveals the benefits of incorporating manifold learning into deep neural networks.

My research is published or under review in the top and leading machine learning conferences and journals, including NIPS, ICML, AISTATS, UAI, ICLR, JMLR and PNAS (Proceeding of National Academy of Science). The application of my machine learning methods have been published in the leading conferences and journals in computer vision and artificial intelligence, including AAAI (Best Poster/Best Presentation Finalist in 2016), IJCAI, CVPR, ECCV (Best Paper Finalist in 2016) and BMVC. The proposed methods enjoy solid theoretical guarantee and demonstrate compelling empirical performance in various computer vision applications by avoiding the curse of dimensionality through low-dimensional structure in data.

My research on learning low-dimensional structure in high-dimensional data include the following works:

- **Subspace Learning**

  High-dimensional data, such as facial image data or gene expression data, often lie in or close to a union of low-dimensional subspaces. Learning the subspace structure is crucial for the success of analytics on such data. We propose ℓ^0-induced Sparse Subspace Clustering (ℓ^0-SSC) and prove that it almost surely recovers the underlying subspaces under randomized models [1], with far-less restrictive assumptions compared to its

ℓ^1 counterpart such as ℓ^1-SSC by Elhamifar et al. ℓ^0-SSC has been widely recognized as an important work in subspace learning which reveals the merit of ℓ^0-induced sparsity in subspace recovery. ℓ^0-SSC is also extended to semi-supervised learning, and it achieves constantly better results than ℓ^1-SSC on extensive image data sets for clustering and semi-supervised learning.

- **Manifold Learning**
  - **Regularized Sparse Graph**
    A sparse graph is a graph which has only a few edges for each vertex, wherein the learned sparse similarity serves as the edge weight. Sparse graph has been demonstrated to be effective for clustering and semi-supervised learning for high-dimensional data. We propose Neighborhood Regularized ℓ^1-Graph (NRℓ^1-Graph) [2] from the perspective of manifold learning with provable guarantee on its optimization. NRℓ^1-Graph constructs sparse graph based on local smoothness along the data manifold and it improves ℓ^1-Graph with convincing empirical results.
  - **Regularized Sparse Coding and Its Fast Encoder: Combining Manifold Learning and Deep Learning**
    Sparse coding represents input signal by a few atoms in a learned dictionary, and it has broad applications to different areas of machine learning and signal processing. We propose Support Regularized Sparse Coding (SRSC) [3] wherein the sparse codes of nearby data in the manifold share the dictionary atoms in accordance with the manifold structure in the high-dimensional data. Moreover, a feed-forward neural network (Deep-SRSC) is proposed to approximate the sparse codes generated by SRSC with significant speedup compared to the traditional iterative optimization algorithm. Experiments on clustering and semi-supervised learning tasks show the advantage of SRSC over canonical sparse coding, as well as the approximation capability of Deep-SRSC.
- **Nonparametric Similarity**
  Similarity-based clustering and classification, such as pairwise clustering, spectral clustering and similarity-based classification, is an important area of machine learning. Similarity measure is crucial for the performance of similarity-based clustering and classification. In addition, learning nonparametric similarity usually circumvents the difficult problem of parameter estimation for parametric models with high-dimensional data. Instead of setting similarity using kernels or K-Nearest-Neighbor graph, the proposed nonparametric similarity is inspired by the generalization analysis of classification in my research [4]. Nonparametric pairwise similarity is induced by the the generalization error bound for unsupervised nonparametric classifiers for the purpose of clustering. It has been extended to discriminative parametric similarity for classification in [5].

I have developed optimization theory for important optimization problems in parallel with my aforementioned research, especially for sparse learning problems. These optimization problems are involved in my research in subspace and manifold learning, and they are important for machine learning and statistics in themselves. I have proposed provable randomized algorithms for efficient optimization on large-scale data. I have also devoted efforts to provable machine learning models by functional analysis and statistical learning theory. In my recent research [5], the generalization bounds for similarity-based classifiers are derived as an extension of our work on nonparametric similarity. Please refer to more details in the next section.

2 Current Accomplishment

My research features both theory and application of learning low-dimensional structure in high-dimensional data. The learned low-dimensional structure includes subspaces, manifolds (through regularized sparse graph and regularized sparse coding), and nonparametric similarity. All these topics are introduced in detail below.

- **Subspace Learning: ℓ^0-Sparse Subspace Learning**
  Subspace learning methods aim to identify data lie in different subspaces. Sparse subspace learning methods are promising in the field of subspace learning and they seek for the sparse codes which satisfy the subspace detection property where nonzero elements of the sparse code of a data point correspond to the data that lie in

the same subspace as that point. When the subspace detection property is satisfied for all the data, a sparse similarity matrix over data can be constructed such that the similarity vanishes between data from different subspaces, which guarantees the satisfactory performance of the subsequent graph-based learning methods such as spectral clustering and label propagation. In contrast to the required assumptions, such as geometric conditions on the subspaces including subspace affinity and subspace incoherence, for most existing sparse subspace clustering methods such as ℓ^1-SSC by Elhamifar et al., we propose ℓ^0-Sparse Subspace Clustering (ℓ^0-SSC) [1] which recovers the subspaces under far less restrictive assumptions. We prove that the subspace detection property can be satisfied by ℓ^0-SSC for arbitrary distinct underlying subspaces almost surely under the randomized models wherein the data in each subspace are distributed at random, and the subspaces are either deterministic or random. We also develop a novel optimization algorithm named Approximate ℓ^0-SSC (Aℓ^0-SSC) that employs proximal gradient descent to obtain a sub-optimal solution to the ℓ^0 sparse approximation problem with theoretical guarantee. ℓ^0-SSC has been extended to semi-supervised learning to form a ℓ^0-Sparse Subspace Learning framework, and extensive experimental results on various data sets including image sets of face and objects demonstrate the superiority of Aℓ^0-SSC for clustering and semi-supervised learning. This work is presented at ECCV 2016 at the optimization session and nominated as Best Paper Finalist among the 11 best paper candidates. We have also proposed dimensionality reduced ℓ^0-SSC wherein ℓ^0-SSC is performed on the dimensionality-reduced data by random projection for improved efficiency [6], as well as provable robust ℓ^0-SSC which guarantees the correctness of ℓ^0-SSC on noisy data [7].

- **Manifold Learning**

  - **Regularized Sparse Graph: Neighborhood Regularized ℓ^1-Graph**

    Sparse graph based methods such as ℓ^1-Graph are proven to be effective on high-dimensional data clustering, but most existing methods construct the sparse graph by performing sparse representation for each data point separately without considering the geometric information of the data. Based on the observation that high-dimensional data always lie in or close to low-dimensional submanifolds, we propose regularized ℓ^1-Graph named Neighborhood Regularized ℓ^1-Graph (NRℓ^1-Graph) which requires vertices corresponding to nearby data in the manifold have similar neighbors in the constructed sparse graph [2]. NRℓ^1-Graph imposes the proposed local smoothness on the neighborhoods by the support distance between the sparse codes in the graph regularization term. The objective function of NRℓ^1-Graph is optimized by proximal gradient descent with theoretical guarantee on the convergence and the obtained sub-optimal solution has bounded gap to the globally optimal solution. NRℓ^1-Graph renders better performance than ℓ^1-Graph for data clustering.

  - **Regularized Sparse Coding and Its Fast Encoder: Combining Manifold Learning and Deep Learning**

    Sparse coding represents input signal by a linear combination of only a few atoms of a learned over-complete dictionary. While broadly applied to various machine learning tasks, the process of obtaining sparse code with fixed dictionary is independent for each data point without considering the manifold structure of the entire data. We propose Support Regularized Sparse Coding (SRSC) [3] which produces sparse codes that account for the manifold structure of the data by encouraging nearby data in the manifold to choose similar dictionary atoms. In this manner, the obtained support regularized sparse codes capture the locally linear structure of the data manifold and possess robustness to data noise. The provable optimization algorithm for SRSC is presented. Moreover, a feed-forward neural network termed Deep Support Regularized Sparse Coding (Deep-SRSC) as a fast encoder for approximating the sparse codes generated by SRSC is designed according to the optimization scheme of SRSC. Extensive experimental results demonstrate the effectiveness of SRSC and Deep-SRSC for clustering and semi-supervised learning tasks.

- **Nonparametric Similarity**

  Similarity-based clustering and classification, including pairwise clustering, spectral clustering and similarity-based methods for classification, is an important area of machine learning. For example, the well-known Support Vector Machine is a special case of similarity-based classification where a positive semi-definite kernel is used as the similarity function. The performance of similarity-based clustering and classification highly depends on the similarity measure between the data. Moreover, learning nonparametric similarity can avoid the difficulty caused by parameter estimation for parametric models with high-dimensional data. We propose nonparametric similarity for similarity-based clustering in [4]. Instead of setting the similarity by kernels or KNN graph, nonparametric similarity is derived by the generalization analysis of classification. Connecting clustering to classification, the general data clustering problem is formulated as a nonparametric unsupervised classification problem wherein the quality of each possible cluster labeling is evaluated by the generalization error of the corresponding classification model, and the optimal cluster labeling corresponds to a classification model with minimum generalization error. Under such setting, the generalization error bounds for the classification models using the plug-in classifier (or the kernel density classifier) and the nearest neighbor classifier are derived via nonparametric generalized kernel density estimator. Both bounds are expressed as sum of nonparametric data similarity for the purpose of clustering, and minimizing such bounds amounts to minimization of the nonparametric similarity between data from different classes. The induced nonparametric similarity leads to better empirical results than kernel similarity. It is also proved that the generalization error bound for the unsupervised plug-in classifier is asymptotically equal to the weighted volume of cluster boundary for Low Density Separation, a widely used criteria for semi-supervised learning and clustering. Our recent work [5] extends the proposed nonparametric similarity to discriminative parametric similarity for classification.

I have developed optimization theory for several important optimization problems involved in my research in learning low-dimensional structure, especially for sparse learning problems. The studied optimization problems are important for the broad machine learning and statistics literature in themselves. Our recent work [8] establishes the conditions under which proximal gradient descent converges to the global optima of the well-known NP-hard ℓ^0 sparse approximation problem, as well as its provable efficient optimization by reducing the dimension of the data via random projection. Our work on pairwise clustering using Markov Random Field (MRF) [9] proposes a novel and efficient message-passing algorithm which reduces the complexity of computing the message along each edge of pairwise MRF from the square of the number of labels to linear complexity, by exploiting the structure of the pairwise potential function. I have also proposed provable randomized optimization algorithms for optimization problems of fundamental importance in my research, and these algorithms are inspired by the established algorithms in the optimization and matrix theory literature, including randomized low-rank matrix approximation and random projection. The proposed randomized algorithms improve the efficiency and scalability of the original deterministic algorithms and guarantee the qualify of the obtained solution. For example, the gap between the sub-optimal solution obtained by randomized low-rank approximation and the globally optimal solution to the original sparse learning problem is presented in my work on NRℓ^1-Graph [2], providing guarantee on the quality of the solution obtained by the proposed randomized algorithm while enjoying the improved efficiency.

I am also very interested in developing provable machine learning models based on functional analysis and statistical learning theory. For example, the asymptotic properties of empirical density estimators such as generalized kernel density estimators are presented in [4]. Moreover, our work on nonparametric similarity has been extended to discriminative similarity parameterized by the parameters of similarity-based classifiers, and the generalization bounds for such classifiers are derived in my recent research [5] by Rademacher complexity and its transductive variant, i.e. transductive Rademacher complexity.

3 Ongoing Research and Future Plan

Base on my previous works, my ongoing and future research include the following topics:

- Robust and efficient optimization algorithms for big data

  The optimization theory for ℓ^0 based sparse regression problem is important for machine learning and statistics, and this problem plays an essential role in my research in subspace learning. We will continue our research on novel and efficient optimization algorithms for robust ℓ^0 based sparse regression which tolerate stochastic or adversarial noise and corruption, and handle missing data. While random projection has been employed to improve the efficiency of ℓ^0 based sparse regression in our previous work, we will study more randomized algorithms, including but not limited to random projection, to improve the efficiency and scalability of optimization algorithms for more general optimization problems, so that these algorithms can handle large-scale data with ease.

- More robust subspace learning on missing data or corrupted data

  It is common that data suffer from missing features or corruption from time to time. We are working on improved ℓ^0-SSC which is robust to missing data or corrupted data with provable correctness of subspace recovery, and this research can benefit from our research on robust optimization on ℓ^0 based sparse regression.

- Understanding deep neural networks

  Our Deep-SRSC is designed according to the conventional iterative optimization algorithm with satisfactory empirical performance. However, there is no theoretical analysis explaining its performance. We are working on the theoretical justification of Deep-SRSC which reveals the reason that Deep-SRSC can approximate regularized sparse codes with only a few layers. We are also working on more joint frameworks of manifold learning and deep learning. Moreover, we plan to extend our analysis to general deep neural networks and attack the open problem on the performance guarantee of general deep neural works. With the stunning success of deep neural networks on various machine learning and artificial intelligence tasks, our analysis is expected to have impact on designing more effective architectures for deep neural networks.

References

[1] Yingzhen Yang, Jiashi Feng, Nebojsa Jojic, Jianchao Yang, and Thomas S. Huang. L0-sparse subspace clustering. In *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II*, pages 731–747, 2016.

[2] Yingzhen Yang, Jiashi Feng, Jiahui Yu, Jianchao Yang, and Thomas S. Huang. Neighborhood regularized l1-graph. In *Proceedings of the Thirty-Third Conference on Uncertainty in Artificial Intelligence, UAI 2017, Sydney, Australia, August 11-15, 2017*, 2017.

[3] Yingzhen Yang, Jiahui Yu, Pushmeet Kohli, Jianchao Yang, and Thomas S. Huang. Support regularized sparse coding and its fast encoder. In *Proceedings of the Fifth International Conference on Learning Representations, (ICLR) 2017, Toulon, France, April 24-26, 2017*, 2017.

[4] Yingzhen Yang, Feng Liang, Shuicheng Yan, Zhangyang Wang, and Thomas S. Huang. On a theory of nonparametric pairwise similarity for clustering: Connecting clustering to classification. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 145–153, 2014.

[5] Yingzhen Yang, Feng Liang, Nebojsa Jojic, Shuicheng Yan, Jiashi Feng, and Thomas S. Huang. Discriminative similarity for clustering and semi-supervised learning. *Submitted to Journal of Machine Learning Research*, 2018.

[6] Yingzhen Yang. Dimensionality reduced ℓ^0 sparse subspace clustering. In *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics*, 2018.

[7] Yingzhen Yang, Jiashi Feng, Nebojsa Jojic, Jianchao Yang, and Thomas S. Huang. Robust and efficient subspace clustering. *Sumibtted to Proceeding of Naiontal Academy of Science*, 2018.

[8] Yingzhen Yang, Jiashi Feng, Nebojsa Jojic, and Thomas S. Huang. On proximal gradient descent for ℓ^0 sparse approximation. *Sumibtted to Annual ACM Symposium on Theory of Computing (STOC)*, 2018.

[9] Yingzhen Yang, Xinqi Chu, Feng Liang, and Thomas S. Huang. Pairwise exemplar clustering. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, July 22-26, 2012, Toronto, Ontario, Canada.*, 2012.