A Third Order based Additional Regularization in Intrinsic Space of the Manifold

Rakesh Kumar Yadav*, Abhishek*, Shekhar Verma*, S Venkatesan*, M. Syafrullah†, Krisna Adiyarta‡
*Department of IT, Indian Institute of Information Technology-Allahbad, Prayagraj, India-211015
†‡Program of Master of Computer Science, Universitas Budi Luhur, Indonesia
Email: *{pcl2014003, rsi2016006, sverma, venkat}@iiita.ac.in
†mohammad.syafrullah@budiluhur.ac.id, ‡krisna.adiyarta@gmail.com

***Abstract*—Second order graph Laplacian regularization has the limitation that the solution remains biased towards a constant which restricts its extrapolation capability. The lack of extrapolation results in poor generalization. An additional penalty factor is needed on the function to avoid its over-fitting on seen unlabeled training instances. The third order derivative based technique identifies the sharp variations in the function and accurately penalizes them to avoid over-fitting. The resultant function leads to a more accurate and generic model that exploits the twist and curvature variations on the manifold. Extensive experiments on synthetic and real-world data set clearly shows that the additional regularization increases accuracy and generic nature of model.**

***Index Terms*—Graph Laplacian, Third order derivative, Regularization, Curvature, Manifold**

I. Introduction

Machine learning techniques [1] aim to extract the underlying common distribution properties from the training data set [2] that can be extended to unseen data instances for the best label prediction. This prediction often suffers from the model over-fitting that deviates the model from obtaining a general characteristic. In many of the real world applications, labeled data is available in sparse while unlabeled data is present in abundance. Semi-supervised learning based classification (SSL) [3]–[5] exploits these combined bundle of labeled as well as unlabeled data [6] that greatly extends the model proficiency. SSL creates a model based on either one of the following predefined assumptions-smoothness, clustering, and manifold. The clustering assumption states that the instances belonging to same cluster must share the similar class label. The classification boundary thus, passes through the regions between clusters where data is sparsely scattered. The manifold assumption inherently assumes that the given high dimensional data actually resides on the much lower dimensional space. Noise, highly correlated attributes and the unknown transformations contribute to the artificially increased dimensions.

Manifold regularization [2], [7]–[10] under SSL framework performs computation on the graph which describes the manifold structure. The nodes of the graph represent the data instances and similarity or affinity between nodes is represented using edges. Consider an undirected graph $G = (V, W)$ where V represents n labeled and unlabeled instances. The edge between data instances x_i and x_j is represented by $w_{ij} \in W$. The similarity metric between nodes is used to extract the intrinsic geometry of the manifold [11], [12]. Graph Laplacian is calculated using $L = D - W$, where $D_{ii} = \sum_{j=1}^{n} w_{ij}$ is a diagonal matrix and L estimates the divergence of the function gradient. Manifold regularization utilizes the abundant unlabeled data to increase the model accuracy and avoid function over-fitting when labeled data alone proves to be insufficient. The manifold regularization accurately penalizes the function so that its transition within similar and non-similar labeled data remains smooth.

Graph Laplacian in past has been proved to be an accurate manifold regularization method but due to its own limitation of extrapolation power, it has been used along with Hessian regularization [13] to achieve better regularization than the graph Laplacian alone. Hessian regularization [14], [15] approximates energy in the neighborhood through the second derivative of the function. It helps in identifying those geodesic deviating functions which is left unpenalized by the graph Laplacian regularization. This higher order co-regularization [16] method successfully discards the effects of noise and high oscillation.

Given d dimensional n data points $X = (x_1, x_2, ..., x_l, x_{l+1}, ..., x_n)$ where, for the sake of simplicity, first l instances are considered labeled and rest $n - l$ are unlabeled data instances. The labels of respective input space data is represented using $Y = (y_1, y_2, ..., y_l)$. As described above, the graph Laplacian on both labeled and unlabeled data is obtained through $L = D - W$ over $G = (V, W)$ where,

$\{w_{ij}\}_{i=1,j=1}^{n}$ represents the similarity between the nodes x_i and x_j. The connectivity between the nodes can be defined on the basis of k-nearest neighbor technique, where x_i and x_j are connected if x_i is in the neighborhood of x_j or vice-versa. The other method for creating a graph is by employing ϵ-radius neighborhood where, x_i and x_j are connected only if $||x_i - x_j||_d^2 < \epsilon$.

A supervised model can be defined as

$$f^* = \underset{f \in H_k}{argmin} \sum_{i=0}^{l} \mathrm{V}(x_i, y_i, f) + \lambda ||f||_A^2 \quad (1)$$

where $\mathrm{V}(x_i, y_i, f)$ represents the loss function for e.g. hinge loss (support vector machine) or square loss (regression least squares) and $\lambda ||f||_A^2$ is the ambient space regularization. In order to extract the intrinsic geometrical information of manifold, one more regularization term needs to be included in the function

$$\begin{aligned} f^* = \underset{f \in H_k}{argmin} & \sum_{i=0}^{l} \mathrm{V}(x_i, y_i, f) + \lambda ||f||_A^2 \\ & + \sum_{i,j=1}^{n} w_{ij} (f(x_i) - f(x_j))^2 \end{aligned} \quad (2)$$

The last term in the equation (2) penalizes the function in the intrinsic space to exploit the hidden geometrical information in the data.

$$\sum_{i,j=1}^{n} w_{ij} (f(x_i) - f(x_j))^2 = \mathbf{f}^T L \mathbf{f}$$

where, L denotes the graph Laplacian of the manifold. Based on the Representer theorem, the optimal function $\mathbf{f}$ can be obtained from $f = \sum_{i=1}^{n} a_i K(x_i, x)$. Here, K denotes the $n \times n$ positive definite kernel gram matrix and a is the function coefficient vector. In this work, we proposed a third order derivative based co-regularization technique to overcome the limitation of second order based manifold regularization method. The rest of this paper is organized in the following sections, section II explains our proposed model, section III represents the extensive experimental results and the last section IV draws the conclusion.

II. Third order derivative and Graph Laplacian associated Regularization Model

A. Proposed Model

The shortcomings of second order based regularization method leads to the need for higher order co-regularization that can neutralize the biasness of the solution of the function. The third order derivative based technique handles the oscillations and twists in the curvature as well as discards the effect of noise in the data. The second order derivative estimates the change in curvature and gives accurate value on a dense manifold. However, as the neighborhood becomes sparse, it dips the model's performance due to its inability to learn the smooth function. Hence, the model prediction remains accurate as long as the learned candidate function is able to take into account all the neighborhood points at each data instance. The Graph Laplacian regularization tends to fail on sparse and rapidly varying manifold. In order to handle such limitation, we propose a third order based co-regularization technique fused in the existing objective function to further accurately penalize the function.

On a smooth Riemannian manifold $\mathcal{M}$, the third order derivative can be defined as

$$f = \int_{\mathcal{M}} ||\bigtriangledown_a \bigtriangledown_b \bigtriangledown_c f||^2_{T_x\mathcal{M} \otimes T_x\mathcal{M} \otimes T_x\mathcal{M}} dV(x) \quad (3)$$

where, $\bigtriangledown_a \bigtriangledown_b \bigtriangledown_c f$ is the third order derivative of f and dV is the volume element which integrates the whole patches of $\mathcal{M}$. Since, due to the presence of large unlabeled data, $\mathcal{M}$ tends to form the single large densely connected structure. In order to extract the underlying hidden information, we need to calculate the tangent space $T_x\mathcal{M}$ at each data instance of X. The third derivative of the underlying curvature at each x_i relies on the manifold $\mathcal{M}$'s properties which is independent of the given coordinate representation. Thus, we need to evaluate an independent coordinate system for each x_i. We assume that given $\mathcal{M}$ holds the basic assumption of local linearity i.e. $\mathcal{M}$ follows euclidean properties around a smaller neighborhood and geodesic being the shortest distance between any two points on that manifold. The norm of the third derivative of $\mathbf{z}$ converges to Frobenius norm of f obtained in the independent coordinates

$$\begin{aligned} ||\bigtriangledown_a \bigtriangledown_b \bigtriangledown_c f||^2_{T_x\mathcal{M} \otimes T_x\mathcal{M} \otimes T_x\mathcal{M}} = \\ \sum_{p,q,r=1}^{u} \left(\frac{\partial^3 f}{\partial x_p \partial x_q \partial x_r} \bigg|_{\mathbf{z}} \right)^2 \end{aligned} \quad (4)$$

After the tangent space $T_x\mathcal{M}$ calculation, the higher order value for curvature can be obtained at each point as

$$\left(\frac{\partial^3 f}{\partial x_p \partial x_q \partial x_r} \bigg|_{\mathbf{z}} \right)^2 = \sum_{i,j=1}^{u} \mathcal{T}^{(j)}_{pqr} f(x_i) \quad (5)$$

The equation 5 has been used for the calculation of the rate of change of the curvature of the given manifold where $\mathcal{T}$ relates objective function with the third order derivative that can be computed by fitting a third order Taylor expansion of f at each point x_i. The Taylor polynomial expansion for third order derivative of f is approximated by

$$\begin{aligned} t_3(x_i) = \sum_{\mathbf{K}=0}^{3} \frac{f^{\mathbf{K}}(0)}{\mathbf{K}!} x_i^{\mathbf{K}} = f(0) + f'(0)x_i + \\ \frac{f''(0)}{2!} x_i^2 + \frac{f'''(0)}{3!} x_i^3 \end{aligned} \quad (6)$$

where, $t_3(x_i)$ is the Taylor's third order derivative, $f(0)$ is a constant value, f', f'', and f''' are first, second and third order derivative of f respectively. The individual derivative terms for first, second and third order derivatives have different degrees of manifold geometrical information extraction scope. The second order manifold regularization lags as the varying and twist in manifold that carries a large chunk of information fails. Thus, the third derivative based regularization in the objective function is needed to handle this concealed information which otherwise remains unexploited by the existing techniques.

The standard least square method is used to fit the high degree polynomial

$$\underset{w\in R^{\mathcal{M}}}{arg\,min} \sum_{i,j=1}^{k} \big((f(x_j) - f(x_i)) - (\phi \mathbf{w})_j\big)^2 \quad (7)$$

The $f(x)$ contains the function value for data samples in neighborhood of x_i and $\phi \in R^{k\times c}$, a design matrix with $c = \frac{5(d^2-3d+4)}{2}$. The monomials ϕ for x_j in neighborhood of x_i is $[x_1, \ldots, x_m, x_1x_1, x_1x_2, \ldots, x_mx_m]$. The Frobenius norm of the higher order function can be obtained from

$$\begin{aligned} || \bigtriangledown_a \bigtriangledown_b \bigtriangledown_c f||^2 &\approx \sum_{p,q,r=1}^{l} \left(\sum_{\alpha}^{k} R^{(i)}_{p,q,r,\alpha} \mathbf{f}_\alpha \right)^2 \\ &= \sum_{\alpha,\beta,\gamma=1}^{k} \mathbf{f}_\alpha \mathbf{f}_\beta \mathbf{f}_\gamma J^{(i)}_{\alpha\beta\gamma} \end{aligned} \quad (8)$$

where, $J^{(i)}_{\alpha\beta\gamma} = \sum_{p,q,r=1}^{l} J^{(i)}_{p,q,r,\alpha} J^{(i)}_{p,q,r,\beta} J^{(i)}_{p,q,r,\gamma}$. Thus, the higher order derivative at any point x_i is the monomial of three degree polynomial fit within that neighborhood.

$$R(f) = \sum_{i=1}^{l} \sum_{p,q,r=1}^{u} \left(\frac{\partial^3 f}{\partial x_p \partial x_q \partial x_r}\bigg|_{\mathbf{z}} \right)^2 =$$

$$\sum_{i=1}^{u} \sum_{\alpha\in_{N_k(x_i)}} \sum_{\beta\in_{N_k(x_i)}} \sum_{\gamma\in_{N_k(x_i)}} \mathbf{f}_\alpha \mathbf{f}_\beta \mathbf{f}_\gamma J^{(i)}_{\alpha\beta\gamma} = (\mathbf{f}^T J \mathbf{f}) \quad (9)$$

The higher order regularization term incorporated in the existing objective function can be modified as

$$\begin{aligned} f^* = \underset{f\in H_k}{argmin} \sum_{i=0}^{l} \mathrm{V}(x_i, y_i, f) + \lambda ||f||^2_A \\ + \gamma ||f||^2_I + \mu(\mathbf{f}^T J \mathbf{f}) \end{aligned} \quad (10)$$

$$\begin{aligned} f^* = \underset{f\in H_k}{argmin} \sum_{i=0}^{l} \mathrm{V}(x_i, y_i, f) + \lambda ||f||^2_A \\ + \gamma ||f||^2_I + \mu ||f||^2_I \end{aligned} \quad (11)$$

B. Model with SVM and RLSC classifier

The support vector machine [17] considers the linear hinge loss for the classification but it cannot appropriately penalize the labeled data points for classification. SVM has been defined that can accommodate the higher order regularization term in the objective function

$$\begin{aligned} f^* = argmin \sum_{i=1}^{l} max(1 - y_i f(x_i), 0) \\ + \lambda \mathbf{a}^T \mathrm{K} \mathbf{a} + \gamma \mathbf{a}^T \mathrm{K}^T \mathrm{K} \mathbf{a} + \mu \mathbf{a}^T \mathrm{K}^T \mathrm{K} \mathbf{a} \end{aligned} \quad (12)$$

where, $K_{n\times n}$ is the kernel gram matrix, and a is the coefficient vector containing values based on second order and third order co-regularization. The higher order regularization method based on RLSC classifier uses the L_2 norm function. The updated objective function is

$$\begin{aligned} f^* = \underset{f\in H_k}{argmin} \sum_{i=1}^{l} ||y_i - f(x_i)||^2_d + \lambda \mathbf{a}^T \mathrm{K} \mathbf{a} \\ + \gamma \mathbf{a}^T \mathrm{K}^T \mathrm{K} \mathbf{a} + \mu \mathbf{a}^T \mathrm{K}^T \mathrm{K} \mathbf{a} \end{aligned} \quad (13)$$

III. Experiment and Results

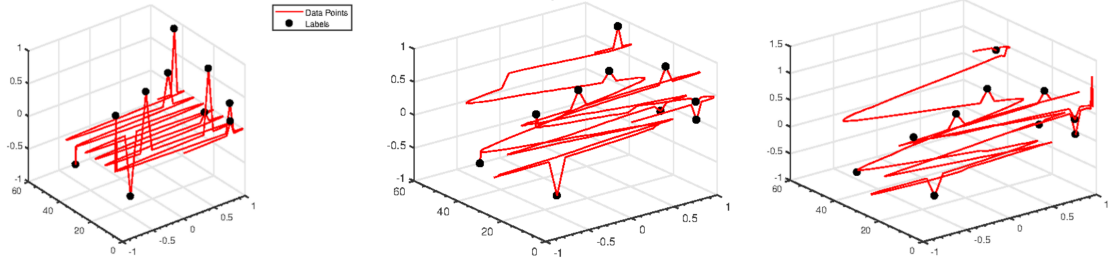
The limitations of existing second order based manifold regularization techniques cannot handle the manifolds with high sparsity and rapidly varying nature with high volume of twist in curvature. The proposed higher order regularization technique overcomes the existing limitations by accounting the rate of change of curvature on the underlying manifold. In this section, we have performed extensive experiments on both synthetic as well as on real world data to validate our newly proposed technique that associates the Graph Laplacian with third order regularization. The performance of our method has been compared with the baseline methods on accuracy metric. The results show that it is able to outperform earlier methods by extracting the additional information concealed in the manifold due to limitations of second order methods.

A. Toy Data set

The proposed technique has been demonstrated on the toy data set that can comprehensible illustrate the better approximate the underlying structure of the 2 d data set. This data set consists of 955 points shaped in a sinusoidal manner. The fig. 2 shows that how well our proposed third order based technique generalize the data points. This generalization of the function is with third order is only possible because it into consideration the rate of change of curvature of the underlying manifold. Comparisons with the existing techniques Graph Laplacian and Hessian regularization has been shown in fig. 1

B. HaLT Data set

This is a large EEG [18] motor data set which contains five BCI paradigms experimental records including HaLT. It is an extension of the 3 state data classic paradigm. It includes left leg, right leg, tongue, left hand, right hand and passive imagery



(a) Sinusoidal Dataset (b) Laplacian (c) Hessian

Fig. 1: Laplacian and Hessian Regularization Technique

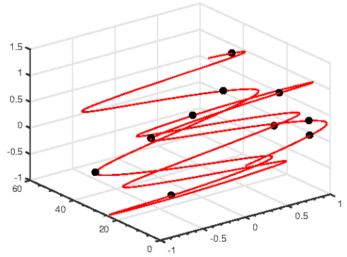
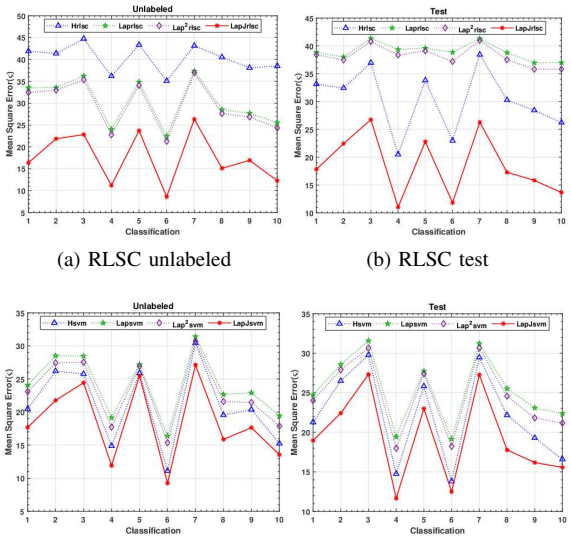


Fig. 2: Proposed Third order based co-regularization



(a) RLSC unlabeled (b) RLSC test

(c) SVM unlabeled (d) SVM test

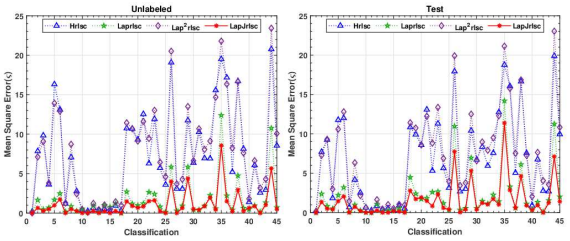
Fig. 3: HaLT data set classification

mental states. Among all the participants, in this experiment, the three HaLT recordings for subject a has been used. In the data collection stage, each movement was shown as image on computer screen for 1 second and simultaneously the respective ECG readings were saved. Each such action consisted of approximately 170 frames of micro-volt data. Based on the certain predefined data markers , each such 170×21 frames extracted and reshaped to a single vector 1×3570. By combining all the frames, the final data set comprises of 2408×3570 matrix. Dimensionality reduction performed with PCA on this high volume data set and reduced to 100 dimensions, retaining the data variance $\approx 90\%$.

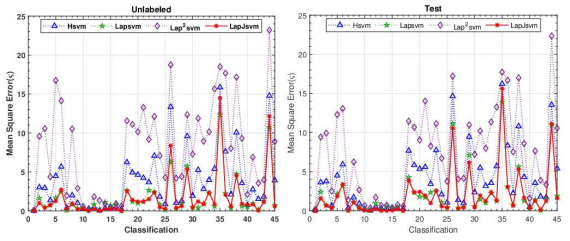
The training and testing data set were generated by randomly dividing action data into two halves. The experiments is performed 10 times such that every time 2 labeled instances is used that generates a highly generic SSL model. Our proposed higher order regularization is compared with other state of art methods Graph Laplacian, Hessian regularization [15] and iterative Laplacian L^m [16], [19] where $(m = 2)$. As show in the figures 3 that proposed higher order regularization technique that outperformed its counterpart with huge margin both for SVM as well as RLSC classification. The minimum and maximum mean squared error (ς) value for Graph Laplacian, Hessian regularization, iterative Laplacian L^m and Higher order regularization are $(36.94, 41.32)$, $(20.48, 38.42)$, $(35.79, 40.99)$, and$(8.60, 23.71)$ respectively. RBF kernel is used to all the method with value of kNN is 6. The value of tuning parameter of λ, γ, and μ is $0.05, 0.005$, and $.004$ respectively is obtained by cross validation. The proposed higher order regularization better extracts the underlying geometry of data manifold that remains unexploited by the second order based techniques which in turn provides a highly generic function.

C. USPS Data set

USPS digit data set [20] is a collection of digits $(0 - 9)$, digitized images of handwritten postal code on postcard in Buffalo NY post office. It is highly enrich data set such that different people have contributed for same digit in their own way(writing style). The main attributes of this data is difference in size , thickness, rotation writing style and instruments. In the experiment each digit has been classified pairwise i.e it is 45 binary classification. Data has total 7291 instances which is reduced to 100 dimensions. Data set is divided into two parts training and testing. 400 instances of each digit is fixed for training and rest instances of each digit is kept for testing. Experiment has been repeated 20 times, on each turn 2 labeled instances are randomly picked for model training. The results (fig 4) shows

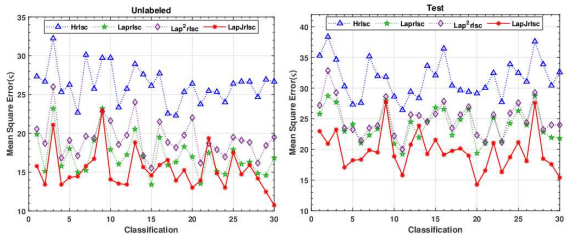


(a) RLSC unlabeled (b) RLSC test

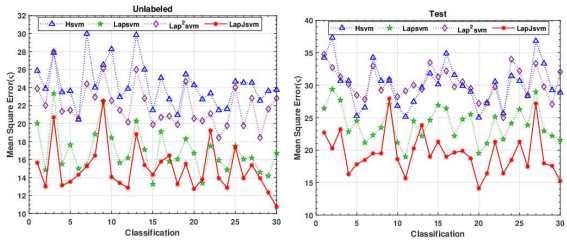


(c) SVM unlabeled (d) SVM test

Fig. 4: USPS data set classification



(a) RLSC unlabeled (b) RLSC test



(c) SVM unlabeled (d) SVM test

Fig. 5: Isolet data set classification

that our higher order proposed regularization performs much better than Hessian regularization and iterative Laplacian but fails to overpass the Graph Laplacian. The minimum and maximum value of mean square error (ς) for Graph Laplacian, Hessian, iterative Laplacian and higher order regularization are $(0.00, 13.91), (00.02, 16.20), (00.17, 22.33),$ and$(0.00, 15.60)$ respectively. RBF kernel is used to all the method with value of kNN is 6. The value of tuning parameter of $\lambda, \gamma,$ and μ is $0.05, 0.005,$ and $.04$ respectively is obtained by cross validation. The higher order based technique here not able to perform much significant and average error value is very much close to the graph Laplacian

D. Isolet Data set

This data generated with the help of 150 individuals who speaks out the names of each English alphabet letter twice. Thus 52 training examples from each speaker. The speakers are clustered together into a set of 30 speakers each, such that five cluster set is formed and termed as Isolet_1,.....Isolet_5. We used Isolet_1 for training and Isolet_5 for testing. Training data has 1560 instances with dimensions 617 whereas testing set has 1559 instances only. Experiment is performed 20 times such that it becomes a 30 binary classification problem. Classification is performed with both SVM and RLSC and our proposed higher based regularization performed better in both cases. It overpasses its counterpart regularization technique by a significant margin. The figure 5 clearly shows the validity of our work such that it classify better for every instance. The minimum and maximum value of mean square error (ς) for Graph Laplacian, Hessian, iterative Laplacian and higher order regularization are $(13.39, 23.20), (22.28, 32.22), (15.51, 25.99),$ and $(10.74, 22.84)$ respectively. RBF kernel is used to all the method with value of kNN is 6. The value of tuning parameter of $\lambda, \gamma,$ and μ is $0.05, 0.005,$ and $.07$ respectively is obtained by cross validation. Thus, the higher order regularization better extracts the underlying geometry of manifold that remains unobserved with the second order based regularization techniques that not take into consideration the twists and sparsity of the underlying manifold.

IV. Conclusion

The proposed higher order manifold co-regularization fused in the objective function along with graph Laplacian over comes the shortcoming of the existing manifold regularization with a significant margin in both SVM and RLSC classification categories. The experimental results on several real world data set illustrate that the higher order co-regularization learns a better generic function by exploiting the rate of change of curvature along with the amount of curvature obtained using graph Laplacian. The higher order derivative based technique identifies the sharp variations in the function and accurately penalizes them to avoid over-fitting. As compared with existing state-of-the-art manifold regularization techniques based on Hessian, graph Laplacian and higher order Laplacian, the proposed method outperforms them by a significant margin.

References

[1] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[2] X. Zhu and A. B. Goldberg, "Introduction to semi-supervised learning," *Synthesis lectures on artificial intelligence and machine learning*, vol. 3, no. 1, pp. 1–130, 2009.

[3] M. Belkin and P. Niyogi, "Semi-supervised learning on riemannian manifolds," *Machine learning*, vol. 56, no. 1-3, pp. 209–239, 2004.

[4] Y. Song, F. Nie, C. Zhang, and S. Xiang, "A unified framework for semi-supervised dimensionality reduction," *Pattern recognition*, vol. 41, no. 9, pp. 2789–2799, 2008.

[5] O. Chapelle, B. Scholkopf, and A. Zien, "Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]," *IEEE Transactions on Neural Networks*, vol. 20, no. 3, pp. 542–542, 2009.

[6] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *Journal of machine learning research*, vol. 7, no. Nov, pp. 2399–2434, 2006.

[7] ——, "On manifold regularization." in *AISTATS*, 2005, p. 1.

[8] M. Belkin, I. Matveeva, and P. Niyogi, "Regularization and semi-supervised learning on large graphs," in *COLT*, vol. 3120. Springer, 2004, pp. 624–638.

[9] S. Ying, Z. Wen, J. Shi, Y. Peng, J. Peng, and H. Qiao, "Manifold preserving: An intrinsic approach for semisupervised distance metric learning," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 7, pp. 2731–2742, 2017.

[10] H. Guo, Y. Mao, and R. Zhang, "Mixup as locally linear out-of-manifold regularization," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 3714–3722.

[11] S. Sun and X. Xie, "Semisupervised support vector machines with tangent space intrinsic manifold regularization," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 9, pp. 1827–1839, 2015.

[12] S. Sun, "Tangent space intrinsic manifold regularization for data representation," in *2013 IEEE China Summit and International Conference on Signal and Information Processing*. IEEE, 2013, pp. 179–183.

[13] H. Liu, W. Liu, D. Tao, and Y. Wang, "Laplacian-hessian regularization for semi-supervised classification," in *Proceedings 2014 IEEE International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*. IEEE, 2014, pp. 203–207.

[14] Z. Guan, J. Peng, and S. Tan, "Manifold ranking using hessian energy." *Int. J. Software and Informatics*, vol. 7, no. 3, pp. 391–405, 2013.

[15] K. I. Kim, F. Steinke, and M. Hein, "Semi-supervised regression using hessian energy with an application to semi-supervised dimensionality reduction," in *Advances in Neural Information Processing Systems*, pp. 979–987.

[16] X. Zhou and M. Belkin, "Semi-supervised learning by higher order regularization," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, 2011, pp. 892–900.

[17] S. Melacci and M. Belkin, "Laplacian support vector machines trained in the primal," *Journal of Machine Learning Research*, vol. 12, no. Mar, pp. 1149–1184, 2011.

[18] M. Kaya, M. K. Binli, E. Ozbay, H. Yanar, and Y. Mishchenko, "A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces," *Scientific data*, vol. 5, 2018.

[19] A. Singh and S. Verma, "Graph Laplacian Regularization With Procrustes Analysis for Sensor Node Localization," *IEEE Sensors Journal*, vol. 17, no. 16, pp. 5367–5376, 2017.

[20] J. J. Hull, "A database for handwritten text recognition research," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 16, no. 5, pp. 550–554, 1994.