

# Subspace Clustering via Ridge Regression

Hongzhi Liu

May 13, 2018

## 1 The Model of Ridge Regression

During the last decade, subspace clustering algorithms have attracted substantial research attention, among which spectral clustering based subspace clustering methods have been popular due to their promising performance such as low-rank representation (LRR) [1] and sparse subspace clustering (SSC) [2].

In order to overcome the limitation of requiring prior knowledge on the usually unknown structures of the errors in practice, we need to learn a kind of ridge regression model [3] which is introduced by Professor Peng.

### 1.1 The Concept of VR3

In the paper of Professor Peng, he adopted the following ridge regression model as Equation (1):

$$J_w = \min_Z \|Y - YZ\|_F^2 + \tau \|Z\|_F^2 \quad (1)$$

where  $J_w$  is a loss function and  $\tau > 0$  is a balancing parameter. Here, the team do not require  $z_{ii} = 0$  because: 1)  $y_i$  is in the intrasubspace of  $y_i$  itself, thus  $z_{ii} \neq 0$  is meaningful; 2)  $\tau > 0$  already excludes potentially trivial solutions such as  $I_n$ ; 3) its effectiveness and efficiency are verified by experiments.

It is noted that Equation (1) starkly differs from the existing subspace clustering models, because 2D data are directly used with projections such that inherent spatial information is retained. After Professor Peng finds the optimal projection matrices, he proposes the following 2D Variance Regularized

Ridge Regression (VR3) model as Equation (2):

$$J_w = \min_{Z, P^T P = I_r, Q^T Q = I_r} \|Y - YZ\|_F^2 + \tau \|Z\|_F^2 + \gamma_1 \text{Tr}(P^T G_P P) + \gamma_2 \text{Tr}(Q^T G_Q Q) \quad (2)$$

where  $G_P$  and  $G_Q$  are defined to be 2D covariance matrices. And  $\gamma_1, \gamma_2 > 0$ , are two balancing parameters. It is seen that by minimizing the above objective function, the projections enable us to capture the most variational information from the 2D data and construct the most expressive self-expression, which lead to a powerful data representation.

### 1.2 The Concept of NVR3

In the last section, I learn about 2D Variance Regularized Ridge Regression (VR3). Then Professor Peng expands their VR3 model to account for nonlinearity in the instance space.

Because nonlinear relationship in the instance space usually exists and is important in real world applications, it is important to take into consideration such nonlinearity. Here, the team aim to consider the nonlinear relationship of the projected data with both projection matrices and present their model of Nonlinear VR3 (NVR3) as Equation (3):

$$J_w = \min_{Z, P^T P = I_r, Q^T Q = I_r} J(Q, P, Z) + \eta_1 \text{Tr}(Z L_P Z^T) + \eta_2 \text{Tr}(Z L_Q Z^T) \quad (3)$$

where  $L_P$  (resp.  $L_Q$ ) is obtained from  $D_P - W_P$  (resp.  $D_Q - W_Q$ ). where  $W = [w_{ij}]$  is the similarity matrix with  $w_{ij}$  being the similarity between  $y_i$  and  $y_j$ , and  $D$  is a diagonal matrix with  $d_{ii} = \sum_j w_{ij}$ . It is seen

that the manifolds are learned adaptively, such that the processes of learning projections, representations, and manifold mutually enhance, thus leading to a powerful data representation.

## 2 Evaluation of VR3 and NVR3

To evaluate the effectiveness of the proposed VR3 and NVR3, several state-of-the-art or recently developed subspace clustering methods are taken as baseline algorithms. For unsupervised learning methods, insensitivity to parameter variations is demanding for enhancing their stability in real world applications.

Professor Peng conducts experiments on Alphadigits data set to illustrate the insensitivities of VR3 and NVR3 to parameter variations. Besides, in Figure 1(a), we fix  $\tau = 600$  to show the performance of VR3 with respect to different combinations of  $\gamma_1$  and  $\gamma_2$ .

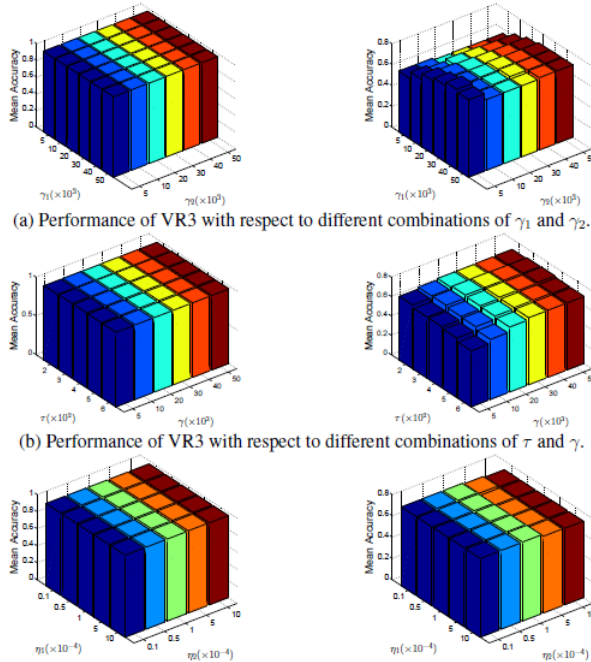


Figure 1: Performance of VR3 and NVR3.  $K = 2$  on the left while  $K = 10$  on the right for (a)-(c).

To demonstrate the effectiveness of this tuning strategy, Professor Peng sets  $\gamma_1 = \gamma_2 = \gamma$  and Figure 1(b) shows the performance of VR3 with various combinations. In fact, if they set  $\eta_1$  and  $\eta_2$  to be zero, then Figure 1(a)-(b) are special cases of NVR3. In Figure 1(c), they show the performance of NVR3 with different combinations of  $\eta_1$  and  $\eta_2$ . The promising performance of the proposed VR3 and NVR3 indicates their potential in real world applications.

## References

- [1] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):171–184, 2010.
- [2] E. Elhamifar and R. Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2765–2781, 2013.
- [3] Chong Peng, Zhao Kang, and Qiang Cheng. Subspace clustering via variance regularized ridge regression. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 682–691, 2017.