LOCALIZED SHRINKAGE COVARIANCE ESTIMATION OF HYPERSPECTRAL IMAGE CLASSIFICATION

Hsiao-Yun Huang[1] *Bor-Chen Kuo*[2] *Jeng-Fu Liu*[3] *Nanping Yang*[4]

[1,3]Department of Statistics and Information Science,
Fu Jen Catholic University, Taipei, Taiwan
[2]Graduate School of Educational Measurement and Statistics
National Taichung University, Taichung, Taiwan
[4]Department of Mathematics,
Fu Jen Catholic University, Taipei, Taiwan

1. INTRODUCTION

The classification of hyperspectral image pixels attracts a lot of attention due to the special high dimension and relatively small sample size data property. Especially when the parametric classifiers, such as linear discriminant classifier and quadratic discriminant classifier, are directly applied, the estimation of the covariance matrix will be a great challenge. Many researches have been devoted to solve this problem. The regularized discriminant analysis (RDA) [1] is a well known classical method. Instead of using regular maximum likelihood covariance estimator, a shrinkage (regularization) estimator of the covariance matrix was employed in RDA. The basic idea of the shrinkage estimation of the covariance matrix Σ is to combine the estimator of the unrestricted model (with many parameters), denoted as U , with the estimator of the restricted model (with few parameters), denoted as T . That is, the shrinkage estimator of Σ , denoted as U^*, is of the form

$$U^* = \lambda T + (1-\lambda)U, \tag{1}$$

where $\lambda \in [0,1]$ represents the shrinkage intensity. In general, the estimator adopted in RDA is the U^* with the maximum likelihood estimator of Σ (denoted as S^{ML}) substituted as U and the scalar covariance estimator $\hat{\sigma}^2 I$ substituted as T [2]. The Leave-One-Out Covariance Estimator (LOOC; [3],[4]), and BLOOC [5] are just the variation version of the U^*.

The performance of the RDA and other parametric classifiers with the regularized covariance estimator depends on the choice of the T and λ simultaneously (usually, the adopted U is S^{ML}). That is, for a selected T , a corresponding shrinkage density λ should be decided. Since only several kinds of T are usually considered [7], how to choose the corresponding λ becomes a crucial problem. Instead of choosing λ by cross-validation as suggested in [1], an analytic result of calculating λ has been derived by Ledoit and Wolf [6]. Several practical considerations and applications of [6] were introduced by [7]. With these new developments, a λ can be determined analytically without requiring computationally expensive procedures. For example, if $U = (u_1, u_2, \cdots, u_p)$ and $T = (t_1, t_2, \cdots, t_p)$, the optimal value is

$$\lambda^* = \frac{\sum_{i=1}^{p} Var(u_i) - Cov(t_i - u_i) - Bias(u_i)E(t_i - u_i)}{\sum_{i=1}^{p} E[(t_i - u_i)^2]} \tag{2}$$

for which the minimum mean square error risk function is achieved. It has been shown in [6] that λ^* always exists and is unique. In practical application, an estimation of λ^* is proposed by [7] as

$$\hat{\lambda}^* = \frac{\sum_{i=1}^{p} \hat{Var}(u_i) - \hat{Cov}(t_i - u_i) - \hat{Bias}(u_i)(t_i - u_i)}{\sum_{i=1}^{p} (t_i - u_i)^2} \quad (3)$$

where all expectations, variances, and covariances in (2) is replaced by their unbiased sample counterparts.

One another view of estimating Σ in high-dimensional case is according to the structure of the estimator. The estimation based on S^{ML} is poor because it imposes too little structure [6]. U^* is designed to improve S^{ML} by imposing another structure T. In this study, we proposed a new estimator of Σ based on the foundation of [6],[7]. This proposed estimator not only imposes structure T analytically, but also introduces the local structure to the estimation. The performance of the proposed estimation of the covariance matrix is compared to the LOOC, BLOOC, and the estimator in RDA via the classification error rate of the corresponding linear and quadratic discriminant classifiers. The results show that the proposed method is an ideal approach for the covariance estimation in hyperspectral image classification.

11. REFERENCES

[1] J. Friedman, “Regularized discriminant analysis,” J. Amer. Stat. Assoc., vol. 84, no. 405, pp. 165–175, Mar. 1989.

[2] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning. New York: Springer-Verlag, 2001.

[3] J. Hoffbeck and D. Landgrebe, “Covariance matrix estimation and classification with limited training data,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 18, no. 7, pp. 763–767, Jul. 1996.

[4] B.-C. Kuo and D. Landgrebe, “A covariance estimator for small samplesize classification problems and its application to feature extraction,” IEEE Trans. Geosci. Remote Sens., vol. 40, no. 4, pp. 814–819, Apr. 2002.

[5] S. Tadjudin and D.A. Landgrebe, Classification of High Dimensional Data with Limited Training Samples, Purdue University, West Lafayette, IN., TR-EE 98-8, pp35-82, April, 1998.

[6] O. Ledoit and M. Wolf, “Improved estimation of the covariance matrix of stock returns with an application to portfolio selection.” J. Empir. Finance 10, pp. 603–621, 2003.

[7] J. Schäfer and K. Strimmer, “A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics,” Statistical Applications in Genetics and Molecular Biology, vol. 4, no. 1, article 32, Nov. 2005.