Kernel Grouped Multivariate Discriminant Analysis for Hyperspectral Image Classification

Mostafa Borhani[✉] and Hassan Ghassemian

Faculty of Electrical and Computer Engineering, Tarbiat Modares University,
Tehran, Iran
{m.borhani,ghassemi}@modares.ac.ir

Abstract. This paper proposes a grouping based technique of multivariate analysis, and it is extended to nonlinear kernel based version for hyperspectral image classification. Grouped multivariate analysis methods are presented in the Euclidean space and dot products are replaced by kernels in Hilbert space for nonlinear dimension reduction and data visualization. We show that the proposed kernel analysis method greatly enhances the classification performance. Experiments on Classification are presented based on Indian Pine real dataset collected from the 224-dimensional AVIRIS hyperspectral sensor, and the performance of proposed approach is investigated. Results show that the Kernel Grouped Multivariate discriminant Analysis (KGMVA) method is generally efficient to improve overall accuracy.

Keywords: Kernel methods · Kernel trick · Multivariate discriminate analysis · Hyperspectral images · Hyperdimentional data analysis · Grouping methods

1 Introduction

Hyperspectral sensors simultaneously capture hundreds of narrow and contiguous spectral images from a wide range of the electromagnetic spectrum, for instance, the AVIRIS hyperspectral sensor [1] has 224 spectral bands ranging from visible light to mid-infrared areas (0.4–2.5 m). Such numerous numbers of images implicatively lead to high dimensionality data, presenting several major challenges in image classification [2–6]. The dimensionality of input space strongly affects performance of many classification methods (e.g., the Hughes phenomenon [7]). This requires the careful design of primitive algorithms that are able to handle hundreds of such spectral images at the same time minimizing the effects from the "curse of dimensionality". Nonlinear methods [8–10], are less sensitive to the data's dimensionality [11] and have already shown superior performance in many machine learning applications. Recently, kernels have a lot of attention in remote-sensed multi/hyperspectral communities [11–16]. However, the full potential of kernels—such as developing customized kernels to integrate a priori domain knowledge—has not been fully explored.

This paper extend traditional linear feature extraction and dimension reduction techniques such as Principal Component Analysis (PCA), Partial Least Squares (PLS),

Orthogonal Partial Least Squares (OPLS), Canonical Correlation Analysis (CCA), NMF (Non-Negative Matrix Factorization) and Entropy Component Analysis (ECA) to kernel nonlinear grouped version. Several extensions (linear and non-linear) to solve common problems in hyper dimensional data analysis were implemented and compared in hyperspectral image classification.

We explore and analyze the most representative MVA approaches, Grouped MVA (GMVA) methods and kernel based discriminative feature reduction manners. We additionally studied recent methods to make kernel GMVA more suitable to real world applications, for hyper dimensional data sets. In such approaches, sparse and semi-supervised learning extensions have been successfully introduced for most of the models. Actually, reduction or selection of features that facilitate classification or regression cuts to the heart of semi-supervised classification. We have completed the panorama with challenging real applications with the classification of land-cover classes.

We continue the paper with an exploring the MVA to the Grouped MVA and then extend the Grouped MVA to the Kernel based Grouped MVA algorithms. Section 3 introduces some simulation of extensions that increase the applicability of Kernel Grouped MVA methods in real applications. Finally, we conclude the paper in Sect. 4 with some discussion.

2   Kernel Grouped Multivariate Analysis

In this section, we first propose the grouping approach and then we extend the linear Canonical Correlation Analysis to kernel based grouped CCA as a sample of kernel based Grouped MVA Methods such as Kernel Grouped Principal Component Analysis (KGPCA), Kernel Grouped Partial Least Squares (KGPLS), Kernel Grouped Orthogonal Partial Least Squares (KGOPLS), and Kernel Grouped Entropy Component Analysis (KGECA). Figure 1 shows the procedure scheme of a simple grouping approach.

For a given a set of observations $\{(x_i, y_i)\}_{i=1}^n$ the grouping algorithm first compute the mean (1) and covariance matrix (2) of entries, where T denotes the transpose of a vector.
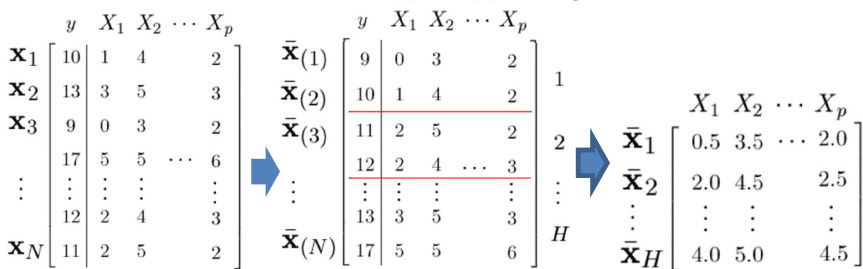

Fig. 1. Procedure scheme of a simple grouping approach

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} \tag{1}$$

$$\hat{\sum}_x = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^T \tag{2}$$

Then extended data set are sorted and collected in H groups. Then again, the procedure leads to compute the mean (3) and weighted covariance matrix (4) of grouped data when n_h is the number of elements in group h and H is the number of groups and N is the total number of elements.

$$\bar{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_i \tag{3}$$

$$\hat{\sum}_W = \frac{n_h}{N} \sum_{h=1}^{H} (\bar{x}_h - \bar{x})(\bar{x}_h - \bar{x})^T \tag{4}$$

The last covariance is explored form the mean of groups and the total mean of elements, like Fisher discriminates analysis. The rest of algorithms are similar the conventional formulation and their extensions to nonlinear kernel based analysis. The use of unbiased covariance formula in (2) and (4) is straight forward.

Canonical Correlation Analysis is usually utilized for two underlying correlated data sets. Consider two iid sets of input data, x_1 and x_2. Classical CCA attempts to find the linear combination of the variables which maximize correlation between the collections. Let

$$y_1 = w_1 x_1 = \sum_j w_{1j} x_{1j} \tag{5}$$

$$y_2 = w_2 x_2 = \sum_j w_{2j} x_{2j} \tag{6}$$

The CCA solves problem of finding values of w_1 and w_2 which maximize the correlation between y_1 and y_2, with constrain the solutions to ensure a finite solution.

Let x_1 have mean μ_1, x_2 have mean μ_2 and $\hat{\sum}_{11}, \hat{\sum}_{22}, \hat{\sum}_{12}$ are denotation of autocovariance of x_1, autocovariance of x_2 and covariance of x_1 and x_2. Then the standard statistical method lies in defining (7). Grouped CCA uses the (4) for computing the covariance of grouped data and K is calculated as (8).

$$K = \hat{\sum}_{11}^{-\frac{1}{2}} \hat{\sum}_{12} \hat{\sum}_{22}^{-\frac{1}{2}} \tag{7}$$

$$K = \sum_{W11}^{\hat{}-\frac{1}{2}} \sum_{W12}^{\hat{}} \sum_{W22}^{\hat{}-\frac{1}{2}} \tag{8}$$

GCCA then performs a Singular Value Decomposition of K to get

$$K = (\alpha_1, \alpha_2, \ldots, \alpha_k) D(\beta_1, \beta_2, \ldots, \beta_k)^T \tag{9}$$

where α_i and β_i are the eigenvectors of Karush–Kuhn–Tucker (KKT) conditions and Tucker-Karush (KTK) conditions respectively and D is the diagonal matrix of eigenvalues.

The first canonical correlation vectors are given by (10) and (11) and in Grouped CCA the canonical correlation vectors are derived from (12) and (13).

$$w_1 = \sum_{11}^{\hat{}-\frac{1}{2}} \alpha_1 \tag{10}$$

$$w_2 = \sum_{22}^{\hat{}-\frac{1}{2}} \beta_1 \tag{11}$$

$$w_1 = \sum_{W11}^{\hat{}-\frac{1}{2}} \alpha_1 \tag{12}$$

$$w_2 = \sum_{W22}^{\hat{}-\frac{1}{2}} \beta_1 \tag{13}$$

As an extension of Grouped CCA, the data were transformed to the feature space by nonlinear kernel methods. Kernel methods are a recent innovation predicated on the methods developed for Support Vector Machines [9, 10]. Support Vector Classification (SVC) performs a nonlinear mapping of the data set into some high dimensional feature space. The most common unsupervised kernel method to date has been Kernel Principal Component Analysis [18, 19]. Consider mapping the input data to a high dimensional (perhaps infinite dimensional) feature space. Now the covariance matrices in Feature space are defined by (14) for i = 1, 2 and covariance matrices of grouped data are by (15) where $\Phi(.)$ is the nonlinear one-to-one and onto function.

$$\sum_{\Phi ij}^{\hat{}} = \frac{1}{N} \sum_{i=1}^{N} \left(\Phi(x_i) - \Phi(\bar{x}) \right) \left(\Phi(x_j) - \Phi(\bar{x}) \right)^T \tag{14}$$

$$\sum_{W\Phi ij}^{\hat{}} = \frac{n_h}{N} \sum_{h=1}^{H} \left(\Phi(\bar{x}_{ih}) - \Phi(\bar{x}) \right) \left(\Phi(\bar{x}_{jh}) - \Phi(\bar{x}) \right)^T \tag{15}$$

However the kernel methods adopt a different approach. w_1 and w_2 exist in the feature space and therefore can be expressed as

$$w_1 = \sum_{i=1}^{2} \sum_{j=1}^{M} \alpha_{ij} \Phi(x_{ij}) \tag{16}$$

$$w_2 = \sum_{i=1}^{2} \sum_{j=1}^{M} \beta_{ij} \Phi(x_{ij}) \qquad (17)$$

where α_i and β_i are the eigenvectors of SVD of $K = \sum_{\hat{\Phi}11}^{-\frac{1}{2}} \sum_{\hat{\Phi}12} \sum_{\hat{\Phi}22}^{-\frac{1}{2}}$ Karush–Kuhn–Tucker conditions and Tucker-Karush conditions respectively for KCCA and α_i and β_i are the eigenvectors of SVD of $K = \sum_{\hat{W}\Phi11}^{-\frac{1}{2}} \sum_{\hat{W}\Phi12} \sum_{\hat{W}\Phi22}^{-\frac{1}{2}}$ KKT and KTK conditions respectively for KGCCA where $K = (\alpha_1, \alpha_2, \ldots, \alpha_k)D(\beta_1, \beta_2, \ldots, \beta_k)^T$ and D is the diagonal matrix of eigenvalues. The rest of Kernel Grouped CCA procedure is similar to KCCA method.

This paper implements several MVA methods such as PCA, PLS, CCA, OPLS, MNF and ECA in linear, kernel and kernel grouped manners. Tables 1, 2 and 3 are summarizing maximization target, Constraints and number of feature of different methods for linear, kernel and kernel grouped approaches where $r(A)$ returns the rank of the matrix A.

Figure 2 shows the projections obtained in the toy problem by linear and modified kernel based MVA methods. Input data was normalized to zero mean and unit variance. Figure 2 shows the features extracted by different MVA methods [20] in an artificial

Table 1. Summary of linear MVA methods

Method	PCA	PLS	CCA	OPLS
Maximize	$u^T C_x u$	$u^T C_{xy} v$	$u^T C_{xy} v$	$u^T C_{xy} C_{xy}^T u$
Constraint	$U^T U = 1$	$U^T U = 1$	$U^T C_x U = 1$	$U^T C_x U = 1$
		$V^T V = 1$	$V^T C_y V = 1$	
# features	$r(X)$	$r(X)$	$r(C_{xy})$	$r(C_{xy})$

Table 2. Summary of kernel MVA methods

Method	KPCA	KPLS	KCCA	KOPLS
Maximize	$\alpha^T T_{\Phi x}^2 \alpha$	$\alpha^T T_{\Phi xy} Y v$	$U^T C_{\Phi xy} V$	$U^T C_{\Phi xy} C_{xy}^T U$
Constraint	$A^T T_{\Phi x}^2 A = 1$	$A^T T_{\Phi x} A = 1$	$A^T T_{\Phi x}^2 A = 1$	$A^T T_{\Phi x}^2 A = 1$
		$V^T V = 1$	$V^T C_{\Phi y} V = 1$	
# features	$r(K_{\Phi x})$	$r(K_{\Phi x})$	$r(K_x Y)$	$r(K_{\Phi x} Y)$

Table 3. Summary of kernel grouped MVA methods

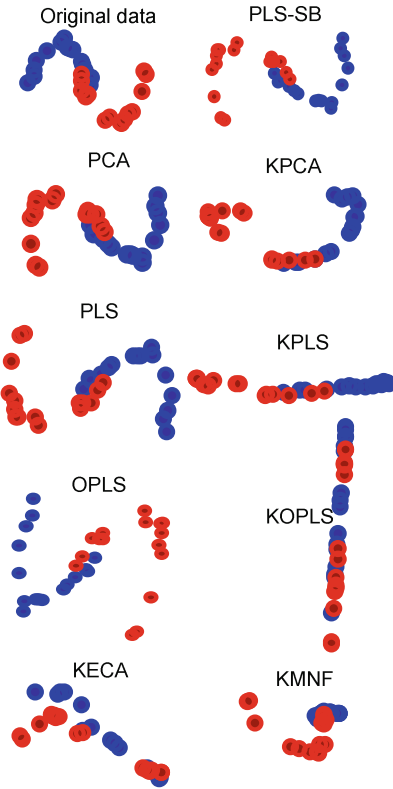| Method	KGPCA	KGPLS	KGCCA	KGOPLS
Maximize	$\alpha^T T_{W\Phi x}^2 \alpha$	$\alpha^T T_{W\Phi xy} Y v$	$U^T C_{W\Phi xy} V$	$U^T C_{W\Phi xy} C_{W\Phi xy}^T U$
Constraint	$A^T T_{W\Phi x}^2 A = 1$	$A^T T_{W\Phi x} A = 1$	$A^T T_{W\Phi x}^2 A = 1$	$A^T T_{W\Phi x}^2 A = 1$
		$V^T V = 1$	$V^T C_{\Phi y} V = 1$	
# features	$r(K_{W\Phi x})$	$r(K_{W\Phi x})$	$r(K_{W\Phi x} Y)$	$r(K_{W\Phi x} Y)$

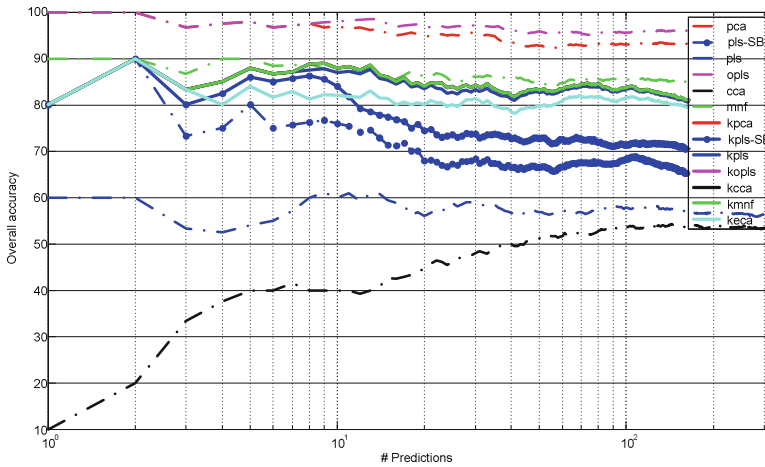**Fig. 2.** Score of various linear MVA, kernel based MVA and kernel grouped MVA methods

Fig. 3. Feature extraction methods: PCA, PLS, OPLS, CCA, MNF, KGPCA, KGPLS, KGOPLS, KGCCA, KGMNF and KGECA, Train Sample = 16
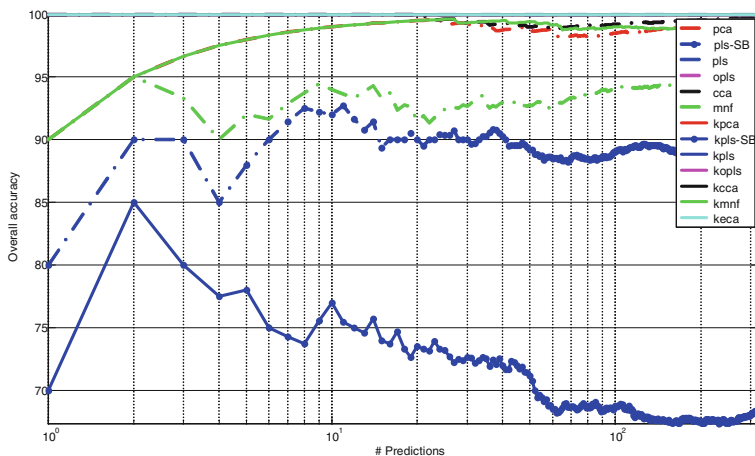
Fig. 4. Feature extraction methods: PCA, PLS, OPLS, CCA, MNF, KGPCA, KGPLS, KGOPLS, KGCCA, KGMNF and KGECA, Train Sample = 144

two-class problem using the RBF kernel. Table 1 provides a summary of the MVA methods and Tables 2 and 3 summarized the kernel MVA and KGMVA methods. For each method it is stated the objective to maximize (First row), constraints for the optimization (second row), and maximum number of features (last row).

3   Experimental Results

Following the kernel grouped dimension reduction schemes proposed in Sect. 2, the performance of the KGMVA methods is compared with a standard SVM with no feature reduction kernel, on AVIRIS dataset. False color composition of the AVIRIS Indian Pines scene and Ground truth-map containing 16 mutually exclusive land-cover classes are showed in Fig. 5.

The AVIRIS hyperspectral dataset is illustrative of the problem of hyperspectral image analysis to determine land use. However the AVIRIS sensor collects nominally 224 bands (or images) of data, four of these contain only zeros and so are discarded, leaving 220 bands in the 92AV3C dataset. At special frequencies, the spectral images are kenned to be adversely affected by atmospheric dihydrogen monoxide absorption. This affects some 20 bands. Each image is of size 145*145 pixels. The dataset was collected over a test site called Indian Pine in north-western Indiana [1]. The database is accompanied by a reference map; signify partial ground truth, whereby pixels are labeled as belonging to one of 16 classes of vegetation or other land types. Not all pixels are so labeled, presumably because they correspond to uninteresting regions or were too arduous to label. Here, we concentrate on the performance of kernel based grouped MVA methods for classification of hyperspectral images. Experimental results are showed in Figs. 3 and 4, for various numbers of train samples and for supervise and unsupervised methods. We use class 2 and 3 for data samples.
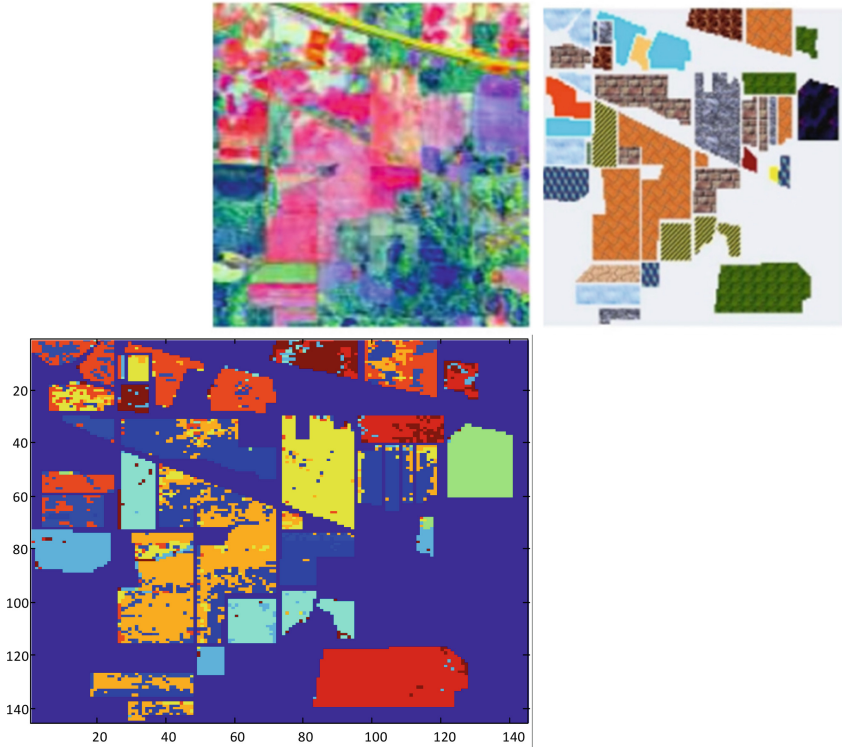
Fig. 5. (Up-Right) False color composition of the AVIRIS Indian Pines scene. (Up-Left) Ground truth-map containing 16 mutually exclusive land-cover classes, (Down-Right) standard SVM, average accuracy = 72.93 % and (Down-Left) SVM with kernel grouped MVA, average accuracy = 79.97, for 64 train samples, 10 classes.

Overall accuracy as a performance measure is depicted v.s. number of prediction for various feature extraction methods such PCA, PLS, OPLS, CCA, MNF, KGPCA, KGPLS, KGOPLS, KGCCA, KGMNF and KGECA. Simulations were repeated for 16 train samples and 144 train samples. Figure 6 shows the average accuracy of different classification approaches, Indiana dataset.

Classification among the major classes can be very difficult [21], which has made the scene a challenging benchmark to validate classification precision of hyperspectral imaging algorithms. Simulations results verified that utilizing the proposed techniques improve the overall accuracy especially kernel grouped CCA in spite of CCA.

4   Discussions and Conclusions

Feature extraction and dimensionality reduction are dominant tasks in many fields of science dealing with signal processing and analysis. This paper provides a kernel based grouped MVA methods. To illustrate the wide applicability of these methods in
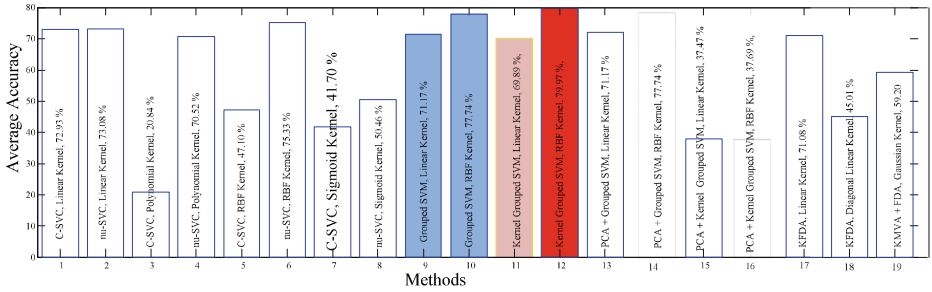
Fig. 6. Average accuracy of different classification approaches, Indiana dataset, 10 classes, 64 train samples. 1. C-SVC, Linear Kernel, 72.93 %, 2. nu-SVC, Linear Kernel, 73.08 %, 3. C-SVC, Polynomial Kernel, 20.84 %, 4. nu-SVC, Polynomial Kernel, 70.52 %, 5. C-SVC, RBF Kernel, 47.10 %, 6. nu-SVC, RBF Kernel, 75.33 %, 7. C-SVC, Sigmoid Kernel, 41.70 %, 8. nu-SVC, Sigmoid Kernel, 50.46 %, 9. Grouped SVM, Linear Kernel, 71.17 %, 10. Grouped SVM, RBF Kernel, 77.74 %, 11. Kernel Grouped SVM, Linear Kernel, 69.89 %, 12. Kernel Grouped SVM, RBF Kernel, 79.97 %, 13. PCA + Grouped SVM, Linear Kernel, 71.17 %, 14. PCA + Grouped SVM, RBF Kernel, 77.74 %, 15. PCA + Kernel Grouped SVM, Linear Kernel, 37.47 %, 16. PCA + Kernel Grouped SVM, RBF Kernel, 37.69 %, 17. KFDA, Linear Kernel, 71.08 %, 18. KFDA, Diagonal Linear Kernel, 45.01 %, 19. KMVA + FDA, Gaussian Kernel, 59.20 %

classification program, we analyze their performance in a benchmark of general available data set, and pay special attention to real applications involving hyperspectral satellite images. In this paper, we have proposed an novel dimension reduction methods for hyperspectral image utilizing kernels and grouping methods. Experimental results showed that, at least for the AVIRIS dataset, the classification performance can be improve to some extent by utilizing either kernel grouped canonical correlation analysis or kernel grouped entropy component analysis. Further work could explore the possibility of localizing grouped of analysis and exploring the algorithms on multiclass datasets. The KGMVA methods were shown to find correlations greater than could be found by linear MVA and also kernel based MVA. However the kernel grouping approach seems to offer a new means of finding such nonlinear and non-stationary correlations and one which is very promising for future research.

References

1. Airborne Visible/Infrared Imaging Spectrometer, AVIRIS. http://aviris.jpl.nasa.gov/
2. Landgrebe, D.: Hyperspectral image data analysis. IEEE Signal Process. Mag. **19**(1), 17–28 (2002). doi:10.1109/79.974718
3. Landgrebe, D.: On information extraction principles for hyperspectral data: a white paper. Technical report, School Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907-1285 (1997). https://engineering.purdue.edu/∼landgreb/whitepaper.pdf
4. Yu, X., Hoff, L.E., Reed, I.S., Chen, A.M., Stotts, L.B.: Automatic target detection and recognition in multiband imagery: a unified ML detection and estimation approach. IEEE Trans. Image Process. **6**(1), 143–156 (1997). doi:10.1109/83.552103

5. Schweizer, S.M., Moura, J.M.F.: Efficient detection in hyperspectral imagery. IEEE Trans. Image Process. **10**(4), 584–597 (2001). doi:10.1109/83.913593

6. Shaw, G., Manolakis, D.: Signal processing for hyperspectral image exploitation. IEEE Signal Process. Mag. **19**(1), 12–16 (2002). doi:10.1109/79.974715

7. Hughes, G.: On the mean accuracy of statistical pattern recognizers. IEEE Trans. Inf. Theor. **14**(1), 55–63 (1968). doi:10.1109/TIT.1968.1054102

8. Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: 5th Annual Workshop on Computational Learning Theory, Pittsburgh, PA, pp. 144–152, (1992). doi:10.1.1.21.3818

9. Cortes, C., Vapnik, V.N.: Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995). doi:10.1023/A:1022627411411

10. Burges, C.: A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Disc. **2**(2), 121–167 (1998). doi:10.1023/A:1009715923555

11. Camps-Valls, G., Bruzzone, L.: Kernel-based methods for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. **43**(6), 1351–1362 (2005). doi:10.1109/TGRS.2005.846154

12. Gualtieri, J., Cromp, R.: Support vector machines for hyperspectral remote sensing classification. In: 27th AIPR Workshop Advances in Computer Assisted Recognition, Washington, DC, pp. 121–132 (1998). doi:10.1.1.27.838

13. Brown, M., Lewis, H.G., Gunn, S.R.: Linear spectral mixture models and support vector machines for remote sensing. IEEE Trans. Geosci. Remote Sens. **38**(5), 2346–2360 (2000). doi:10.1109/36.868891

14. Roli, F., Fumera, G., Serpico, S.B. (ed.) Support vector machines for remote-sensing image classification. In: Proceedings of SPIE Image and Signal Processing for Remote Sensing VI, vol. 4170, pp. 160–166 (2001). doi:10.1.1.11.5830

15. Lennon, M., Mercier, G., Hubert-Moy, L.: Classification of hyperspectral images with nonlinear filtering and support vector machines. In: IEEE International Geoscience and Remote Sensing Symposium 2002, IGARSS'02, 24–28 June 2002, vol. 3, pp. 1670–1672 (2002). doi:10.1109/IGARSS.2002.1026216

16. Melgani, F., Bruzzone, L.: Classification of hyperspectral remote sensing images with support vector machines. IEEE Trans. Geosci. Remote Sens. **42**(8), 1778–1790 (2004). doi:10.1109/TGRS.2004.831865

17. Mardia, K.V., Kent, J.T., Bibby, J.M.: Multivariate Analysis, 1st edn. Academic Press, New York (1980). ISBN 10: 0124712525, 13: 978-0124712522

18. Scholokopf, B., Smola, A., Muller, K.-R.: Nonlinear component analysis as a kernel eigenvalue problem. Technical report 44, Max Planck Institute fur biologische Kybernetik, December 1996. doi:10.1.1.29.1366

19. Scholokopf, B., Smola, A., Muller, K.-R.: Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput. **10**, 1299–1319 (1998)

20. Arenas-Garcia, J., Petersen, K., Camps-Valls, G., Hansen, L.K.: Kernel multivariate analysis framework for supervised subspace learning: a tutorial on linear and kernel multivariate methods. IEEE Signal Process. Mag. **30**(4), 16–29 (2013). doi:10.1109/MSP.2013.2250591

21. M. Borhani, H. Ghassemian, Novel Spatial Approaches for Classification of Hyperspectral Remotely Sensed Landscapes, Symposium on Artificial Intelligence and Signal Processing, December 2013