

THE INVESTIGATION OF SENSITIVITY OF SVM CLASSIFIER RESPECT TO THE NUMBER OF FEATURES AND THE NUMBER OF TRAINING SAMPLES

M. Imani, H. Ghassemian

Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran
maryam.imani@modares.ac.ir
ghassemi@modares.ac.ir

KEY WORDS: Support vector machine, Hyperspectral images, the number of features, the number of training samples

ABSTRACT:

Supervised classification of hyperspectral images is a difficult task due to the unbalance between the high dimensionality of the data and the limited availability of labeled training samples. Recently support vector machine (SVM), has received considerable attention for classifying high dimensional data and is applied successfully for classification of hyperspectral images because it discriminates classes by a geometrical criterion not by statistical criteria. In this paper, we investigate sensitivity of SVM classifier respect to two factors. The first factor is the dimensional of data (the number of features) and the second factor is the number of training samples. We evaluate the effect of these factors on the performance of classification in the point of view both accuracy and reliability. Experiments are carried out on the three different common used hyperspectral datasets, Indian pines, Pavia University and Salinas.

1. INTRODUCTION

By developing hyperspectral imaging technology, it is possible to simultaneously capture image with hundreds of contiguous narrow spectral bands. Recently SVM is applied successfully for classification of hyperspectral images as a non-parametric classifier (Gualtieri, 1999, Pal, 2010, Tarabalka, 2010, Braun, 2012). Since discrimination of classes is based on geometrical criteria not by statistical criteria, they can work even with limited training sample size hence it can overcome the Hughes phenomenon. SVM works based on finding an optimum hyperplane that maximized the margin between two classes. If training data are not separated linearly, a kernel method is used to map data to higher dimension space where data are separated linearly (Moustakidis, 2012, Pal, 2012). In this paper, we investigate the sensitivity of SVM classifier respect to the number of features and the number of training samples. We do experiments using three common used hyperspectral images in different number of features (using PCA) and with different number of training samples.

This paper organized as follows: a brief description of SVM is given in section 2. The experimental results and evaluation of sensitivity of SVM classifier respect to the number of features and training sample size is represented in section 3. Finally this paper is included in section 4.

2. A DESCRIPTION OF SVM

Let us assume that training samples are $(\mathbf{x}_i)_{d \times 1}$ ($i = 1, \dots, N$) and the labels of them are $y_i \in \{-1, +1\}$. If two classes are linearly separable, then we can find one hyperplane that separates the two classes. The discriminant function is defined as follows:

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b \quad (1)$$

where $\mathbf{w}_{d \times 1}$ is a vector normal to hyperspectral and b is bias. In order to find the separable hyperplane, \mathbf{w} and b should be estimated so that:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \quad i = 1, \dots, N \quad (2)$$

The SVM approach finds the hyperplane that maximize margin (the maximum distance between the closet training samples) between two classes. This margin is equal to $\frac{1}{\|\mathbf{w}\|}$. The optimal hyperplane can be found by solution of following optimization problem:

$$\min \frac{1}{2} \|\mathbf{w}\|^2$$

subject to $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \quad i = 1, \dots, N \quad (3)$

using a Lagrangian formulation, the discriminant function associated with the optimal hyperplane is obtained by:

$$f(x) = \sum_{i \in T} \alpha_i y_i (\mathbf{x}_i \cdot \mathbf{x}) + b \quad (4)$$

Where α_i denote the Lagrange multipliers and T is the subset of training samples corresponding to the nonzero Lagrange multipliers. These training samples are called support vectors. If two classes are nonlinear separable, we can map the data through a proper mapping $\Phi(\cdot)$ into a higher dimensional feature space. By replacing mapping function $\Phi(\mathbf{x}_i)$ instead of \mathbf{x}_i and inner product $(\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}))$ by proper kernel functions, $K(\mathbf{x}_i, \mathbf{x})$, (Camps, 2005) the nonlinear SVM is produced and expressed as follows:

$$f(x) = \sum_{i \in T} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \quad (5)$$

3. EXPERIMENTAL RESULTS AND EVALUATION OF SENSITIVITY OF SVM

3.1 Datasets and Evaluation Measures

Three different datasets are used for experiments in this section. The first hyperspectral data is Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Indian pines image. This agriculture image has spatial dimension 145×145 and contains 16. The AVIRIS sensor generates 220 bands that we reduced the number of them to 190 by removing 30 noisy bands. The second used dataset, university of Pavia, is acquired by Reflective Optics System Imaging Spectrometer (ROSIS). This urban image contains 103 bands and 9 classes and is consist of 610×340 pixels. Salinas scene that collected by the 224 band AVIRIS sensor over Salinas valley is the third used dataset that 20 absorption bands of it is discarded and 204 reminded bands are used in the our experiments. Salinas image is a 512×217 image that contains 16 classes. The number of samples in each dataset is given in table 1. Also the ground truth map (GTM) of Indian pines is shown in Figure 1 and GTM of two other datasets are seen in Figure 2.

The used measures for evaluation of performance of classifiers are average class accuracy and average class reliability. Accuracy and reliability for each class defined as follows: accuracy is the number of test samples that are correctly classified divided to the overall test samples and reliability is number of test samples that are correctly classified divided to the overall samples that are labeled as this class.

Number of samples in each class	Indiana	Pavia university	Salinas
Class1	54.00	6631.00	2009.00
Class2	1434.00	18649.00	3726.00
Class3	834.00	2099.00	1976.00
Class4	234.00	3064.00	1394.00
Class5	497.00	1345.00	2678.00
Class6	747.00	5029.00	3959.00
Class7	26.00	1330.00	3579.00
Class8	489.00	3682.00	11271.00
Class9	20.00	947.00	6203.00
Class10	968.00		3278.00
Class11	2468.00		1068.00
Class12	614.00		1927.00
Class13	212.00		916.00
Class14	1294.00		1070.00
Class15	380.00		7268.00
Class16	95.00		1807.00

Table 1. the number of samples in each dataset



Figure 1. Ground truth map of Indian pines dataset

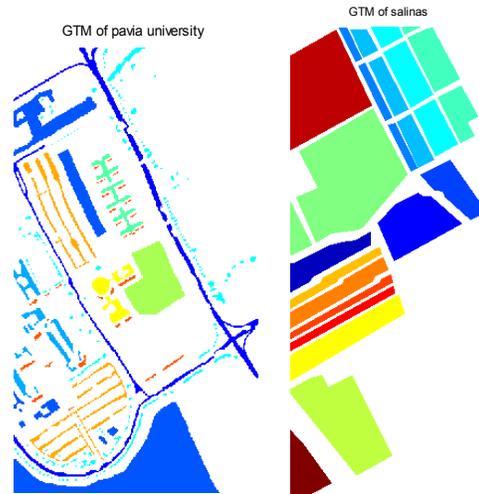


Figure 2. GTM of Pavia university and Salinas datasets

3.2 Experiments and Evaluation

The first experiments are for investigation of sensitivity of SVM classifier respect to the number of features. Feature reduction is done using PCA. The number of used training samples per class (N_t) in investigation of the number of feature is fixed and considered equal to 10. The obtained results are shown in Figure 3. It can be seen that with increasing the number of features, the performance of classifier, both accuracy and reliability, is improved generally to a number and after that reminds in that same value without any increase. The obtained class maps for three hyperspectral images are shown in Figures 4-6. The second experiments are done for investigation of the number of training samples in classification performance. In these experiments, the number of features (N_f) is fixed and equal to 16, 20 and 15 for Indiana, Pavia and Salinas datasets respectively. The results are shown in Figure 7. As we can see, with increasing the number of used training samples, the performance of classifier is improved considerably to a number and after that, this increase is not notable. The obtained class maps for these experiments are shown in Figures 8-10.

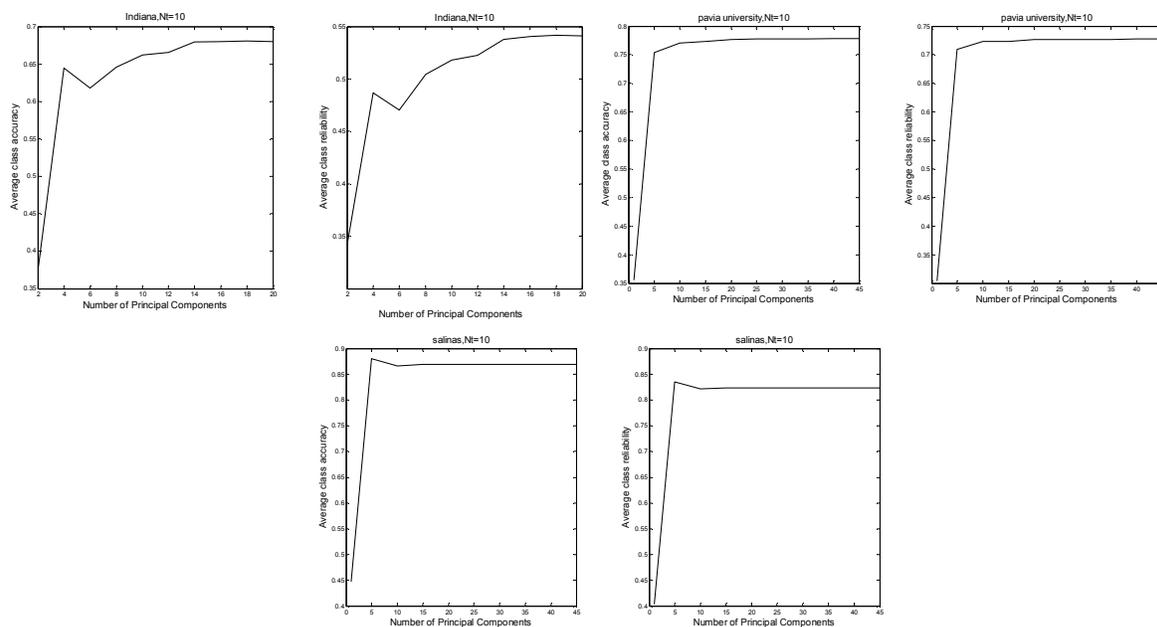


Figure 3. The average class accuracy and average reliability accuracy versus the number of principal components

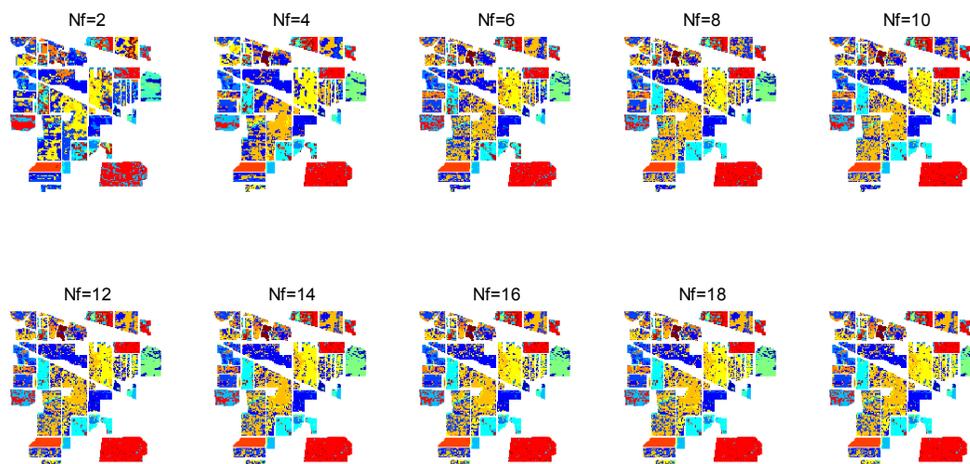


Figure 4. The obtained class maps with $N_t = 10$ and different N_f for Indiana dataset

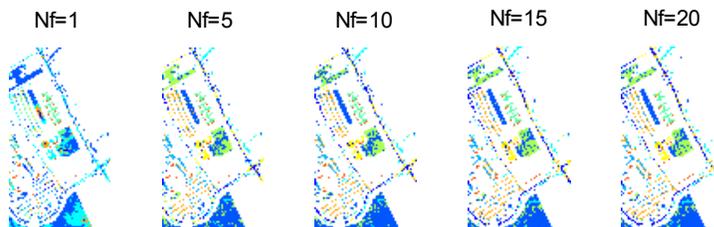


Figure 5. The obtained class maps with $N_t = 10$ and different N_f for Pavia university dataset

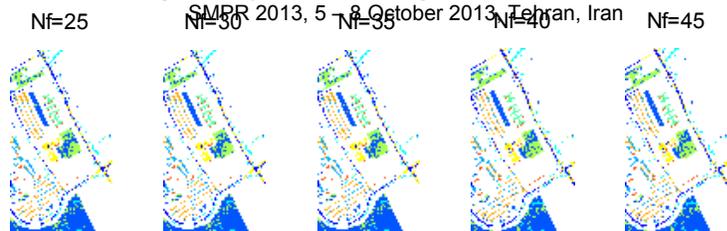


Figure 5 (continue). The obtained class maps with $N_t = 10$ and different N_f for Pavia university dataset

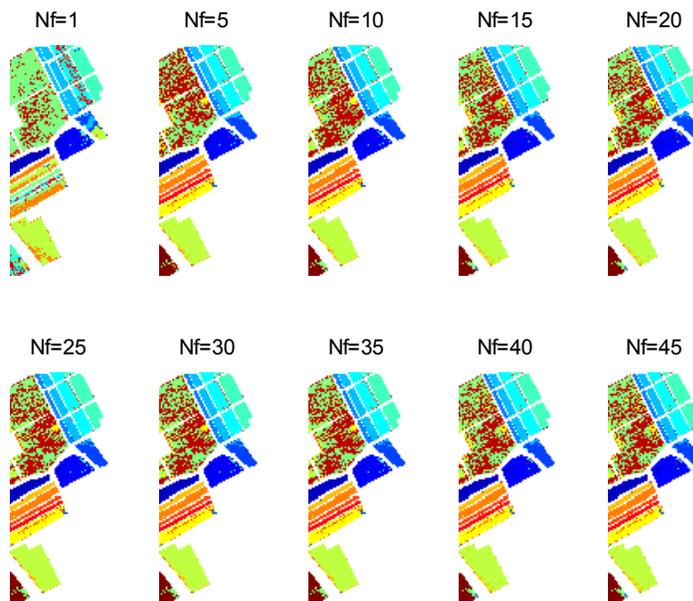


Figure 6. The obtained class maps with $N_t = 10$ and different N_f for Salinas dataset

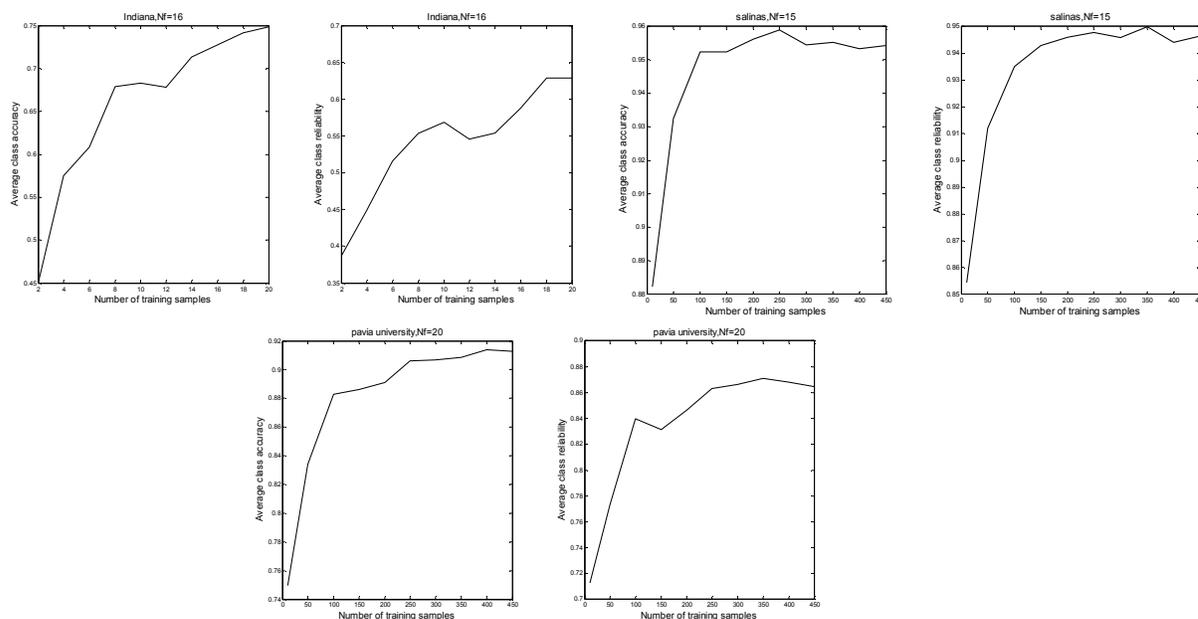


Figure 7. The average class accuracy and average reliability accuracy versus the number of training samples

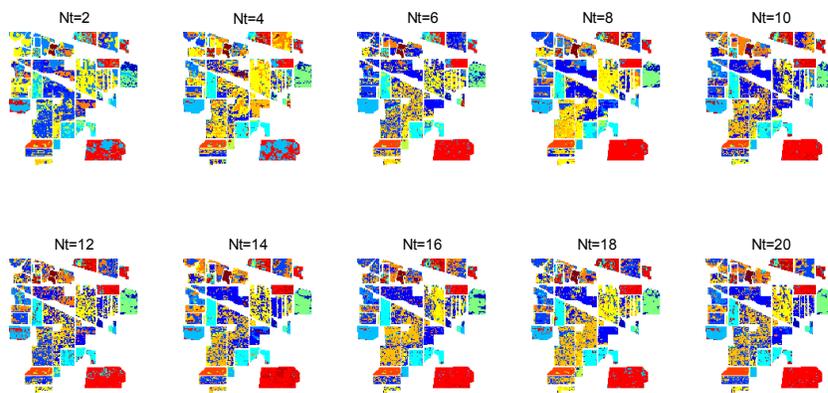


Figure 8. The obtained class maps with $N_f = 16$ and different N_t for Indiana dataset

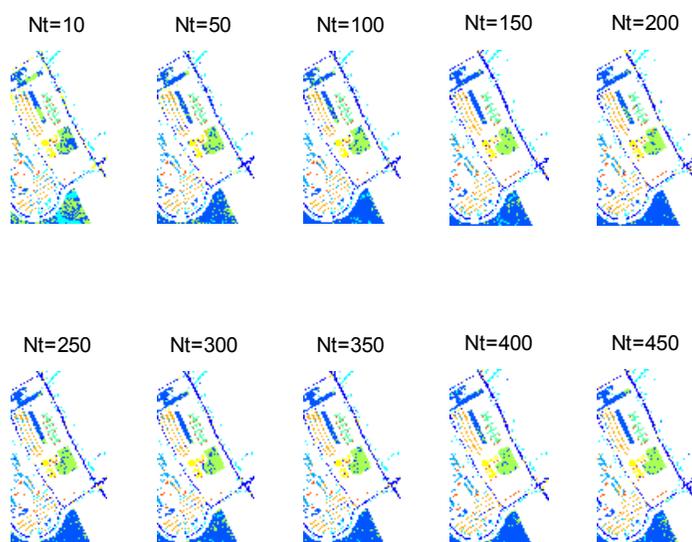


Figure 9. The obtained class maps with $N_f = 20$ and different N_t for Pavia university dataset

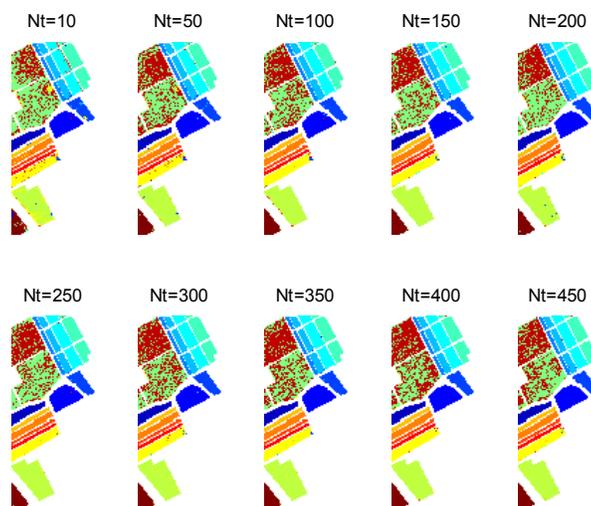


Figure 10. The obtained class maps with $N_f = 15$ and different N_t for Salinas dataset

4. CONCLUSION

In this paper, we evaluate the effect of the number of features and training samples on the performance of SVM classifier. We used PCA for feature reduction. With increase the number of features, the performance is improved generally to a number and after that increasing of the number of feature has no effect on the performance of classification. Also, increasing of the number of used training samples is not considerable with more increase after a number. These results obtained for three common used hyperspectral images.

REFERENCES

- Braun, A.C., 2012, Weidner, U., Hinz, S. Classification in High-Dimensional Feature Spaces. Assessment Using SVM, IVM and RVM With Focus on Simulated EnMAP Data. *Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2),pp. 436 – 443 (April 2012).
- Camps-Valls G., 2005 and Bruzzone L. Kernel-based methods for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.*, 43(6), pp. 1351–1362 (Jun. 2005).
- Gualtieri J. A., 1999, Chettri S. R., Cromp R. F., and Johnson L. F. Support vector machines classifiers as applied to AVIRIS data. in *Proc. Summaries 8th JPL Airborne Earth Sci. Workshop: JPL Publication 99- 17*, pp. 217–227 (Feb. 1999).
- Moustakidis, S., 2012, Mallinis, G., Koutsias, N., Theocharis, J.B.; Petridis, V.SVM-Based Fuzzy Decision Trees for Classification of High Spatial Resolution Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.*, 50(1), pp. 149 -169 (Jan. 2012).
- Pal, M., 2010, Foody, G.M. Feature Selection for Classification of Hyperspectral Data by SVM. *IEEE Trans. Geosci. Remote Sens.*, 48(5), pp. 2297 – 2307 (May 2010).
- Pal, M., 2012, Foody, G.M. Evaluation of SVM, RVM and SMLR for Accurate Image Classification With Limited Ground Data. *Selected Topics in Applied Earth Observations and Remote Sensing*, 5(5),pp. 1344 – 1355 (Oct. 2012).
- Tarabalka, Y., 2010, Fauvel, M., Chanussot, J.,and Benediktsson, J.A. SVM- and MRF-Based Method for Accurate Classification of Hyperspectral Images. *IEEE Trans. Geosci. Remote Sensing Lett*, 7(4), pp. 736 – 740 (Oct. 2010).