SUPPORT VECTOR MACHINES FOR CONSERVATION RESERVE PROGRAM (CRP) MAPPING AND COMPLIANCE MONITORING

by

Ginto Cherian

Bachelor of Technology

Mahatma Gandhi University

Kerala, India

2001

Submitted to the Faculty of the Graduate College of the

Oklahoma State University in partial fulfillment

of the requirements for the Degree of

MASTER OF SCIENCE

December, 2004

SUPPORT VECTOR MACHINES FOR CONSERVATION RESERVE PROGRAM (CRP) MAPPING AND COMPLIANCE MONITORING

By

Ginto Cherian

Thesis Approved:

Dr. Guoliang Fan (Thesis Advisor)

Dr. Jong-Moon Chung

Dr. Gary G. Yen

Dr. Mahesh Rao

Dr. Gordon Emslie (Dean of the Graduate College)

ACKNOWLEDGMENTS

I would like to express my most sincere thanks to my advisor, Dr. Guoliang Fan, for his guidance, support, and encouragement throughout my MS study. He has imparted not only technical knowledge, but also a rigorous attitude towards research. I would also like to acknowledge my colleague Xiaomu Song who participated and helped me in this work. I would like to express my thanks to Dr. Mahesh Rao for providing the data and help on Remote Sensing . Many thanks to my other committee members: Dr. Jong-Moon Chung and Dr. Gary G. Yen for their constructive suggestions and comments.

Most of all, I want to express my deepest thanks to my parents, Olevakunnel Varkey Cherian and Mary Cherian. I cannot thank them enough for their everlasting love, support, and understanding; those are the most precious gifts in my life. My M.S. work would be impossible without their strongest support. They made a tremendous effort to create conditions conductive to good education. Finally, I am also grateful to my brother, Anish Cherian, whose support is valued.

This work was supported by research grants from Oklahoma NASA EPSCOR and OSU Environmental Institutes Water Research Center.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

This thesis deals with the application of Support Vector Machines (SVM) to specific problems in remote sensing and the adaptations required for solving them. The specific problems are mapping and compliance monitoring of US Department of Agriculture's (USDA's) CRP.

This chapter briefly introduces USDA's CRP program and the problems that we intend to study. The major algorithms that are used in this work are discussed in Chapter 2. Datasets and the study area used in our simulations are also presented.

1.1 USDA's Conservation Reserve Program (CRP)

The Conservation Reserve Program (CRP) is a voluntary program for agricultural landowners. It provides annual rental payments for the establishment of long term ecologically-beneficial covers on eligible lands. The program is administered by the Commodity Credit Corporation through the Farm Service Agency both of which are subsidiaries of the USDA. Information about the CRP program is available at the CRP website [1].

Overall speaking, the CRP is a long-term program which aims to improve soil, water and wildlife resources. Under the CRP contract, farmers are encouraged to plant long-term native plant species (mostly grasses) on agricultural lands for a period of 10-15 years. These CRP tracts (e.g. Figure 1.2) have to be maintained according to CRP contract stipulations, which specify that the land cannot be used for

commercial purposes except for haying or grazing during weather-related emergencies. In return annual rental payments are made to the farmers by USDA. The CRP program was established by the Congress in 1985. During the year 2003, 34,110,536 acres were enrolled in the program. The rental payments made amounted to $1.673 billion.

Figure 1.1: Example of CRP lands from [1].

The eligibility of the land for enrollment in the CRP program is based on various factors given below,

- A cropland that has been planted or considered planted for 4 of the last 6 years.
- A cropland that is cultivable.
- A cropland must have a weighted average Erosion Index of 8 or greater.

Offers received for CRP enrollment are evaluated and ranked based on an Environmental Benefit Index (EBI). EBI usage ensures that only the most environmentally sensitive lands are selected. The EBI factors include,

- Wildlife habitat benefits.
- Water quality benefits due to reduced erosion, runoff and leaching.
- On-farm benefits from reduced erosion.
- Air quality benefits from reduced wind erosion.
- Cost.

Lands enrolled in the CRP program have to adhere to contract stipulations. Mainly the land has to be planted with native vegetation usually native grasses. In some cases even trees are allowed. Native vegetation is preferred because one objective is to improve wildlife habitats. Also farming is not allowed during the contract period. Haying or grazing on enrolled lands is not allowed expect during weather related emergencies like drought.

1.2 Remote Sensing for CRP related research

1.2.1 CRP Compliance Monitoring

Currently, USDA is faced with the problem of farmers not maintaining CRP tracts according to contract stipulations. So there is a need to make sure that enrolled CRP lands are maintained properly i.e. compliance monitoring. Current methods for CRP compliance monitoring involve intensive manual inspection of aerial photographs which is time-consuming and costly. USDAs Common Land Unit (CLU) data [15] which is used for general compliance issues, is generated from aerial photographs created by the National Agricultural Imagery Program (NAIP) with a resolution about $1m \times 1m$, which are updated every 1-2 years and may not be

very efficient for CRP compliance monitoring on a large scale. In addition, existing CRP reference data obtained from USDAs Natural Resource Conservation Service (NRCS) is not very accurate or up-to-date for management purposes. Furthermore, random field inspections are costly affairs. There is an urgent need of an automatic compliance monitoring method which can examine CRP tracts more efficiently and promptly with minimum human involvement. Little research has been done in this area.

1.2.2 CRP Mapping

Existing CRP reference data provided by NRCS has some errors and is out-of-date. Usually, major errors in the present CRP reference data are the mis-location and/or misalignment of CRP tracts. This is due to the fact that the reference data is considerably old and it is possible that there have been new CRP enrollments or that old enrollments have expired and returned to agriculture. Current CRP maps are developed based on information provided by farmers upon enrollment into the program and by manual delineation of aerial photographs. So it is necessary to update CRP maps regularly and automatically. Past research in CRP mapping is summarized in the succeeding paragraphs.

An approach for accurate CRP mapping based on multi-seasonal and multi-year Landsat TM imagery is discussed in[8] and [9]. An unsupervised classification was performed first to create crop and grass maps. Then after labelling these clusters manually, the CRP tracts were extracted by a post-classification comparison technique, where the areas with changed cover types can be detected. Although high classification accuracy had been achieved by this approach, the dependency on intensive human skill and labor might limit its efficiency and effectiveness in practical applications to large areas.

In [33], an automated accurate classifier of multi-source geospatial data was developed where the geospatial data consists of multi-temporal Landsat imagery,

ancillary geographic information system (GIS) data and other derived features are involved. Two machine learning approaches, i.e., decision tree classifier (DTC) and SVM, were implemented as multi-source geospatial data classifiers.

1.3 Machine Learning Approach for Remote Sensing

Machine learning is the ability of a computer algorithm to recognize patterns that have occurred repeatedly and to improve its performance based on past experiences. In remote sensing classification, the statistical maximum likelihood (ML) classifier is a widely used tool for land cover classification of multi-spectral imagery. Each land cover class is assumed to be a unimodal normal distribution which is not usually true due to the random nature of remote sensing data. Although the ML method is robust with some deviations in modeling, and the Gaussian mixture model is applied to better approximate the true distribution, the intensive computation load makes it impractical for real applications. In addition, there are some potential relationships among different features, simple statistical models cannot capture those relationships efficiently. In recent years, some machine learning approaches have been successfully applied to remote sensing data [33], [31], [3], [11], [7], [17], [21], [4], [16], [27], [33], [10], [24] and [13]. These approaches are efficient, robust and flexible without any requirement of statistical modeling. As there are currently a huge number of machine learning methods and many of these methods have been applied to problems in Remote Sensing; we will be discussing a few of the important methods.

The Decision Tree Classifier (DTC) solves a complex classification problem by dividing it into a set of simpler classification issues. The DTC has shown advantages in real remote sensing applications for more than ten years [33], [31], [3], [11] and [7]. However, considering the complexity of feature space, the trained DTC might have the over-fitting problem with poor generalization performance. Pruning could mitigate the effect of this problem, but cannot guarantee an optimal solution.

The SVM has recently enjoyed wide usage for remote sensing applications. SVM separates different classes by finding optimal classification hyperplanes from training data, leading to better generalization capability compared with other methods. Recent research on SVM in Remote Sensing applications have shown impressive classification results [17], [21], [4], [16], [27], [33], [10], [24] and [13]. More elaboration on SVM for Landsat TM image classification is given in Section (2.5).

1.4 Experimental Setup

The study area that we have chosen is Texas County (Figure 1.2), Oklahoma. As of October 2003 Texas county had 217,802 acres out of the total 1,036,441 acres enrolled in Oklahoma. Here different grass species are grown on the CRP tracts. This allows us the opportunity to analyze the CRP plots better as plots will have be less variation due to changes in topography and other factors.

In our work we are using Landsat TM (Thematic Mapper) images obtained for February, 2000 and June, 2000 (so as to get information for both winter and summer) covering Texas County, Oklahoma. Hence this data set is multi-temporal. Each pixel of this image covers an area of $30m \times 30m$. Landsat TM generated imagery covers seven spectral bands viz. Band 1 (Blue), Band 2 (Green), Band 3 (Red), Band 4 (Near infrared(IR)), Band 5 (Mid IR), Band 6 (Thermal IR) and Band 7 (Mid IR).

We also have CRP Reference data obtained from Natural Resources Conservation Service which is used as the ground data. This data shows the CRP tracts as well as the cover types in each CRP tract as per the information given by the farmer when his/her land is enrolled in the CRP program.

Feature Extraction is adapted from [33]. Each pattern(representing a single pixel) composed of totally 38 layers generated solely from the Landsat TM images. The first 10 layers of each pattern consist of Landsat TM bands from each TM image. Bands 1 (i.e. blue band is prone to haze) and 6 (i.e. thermal band which has

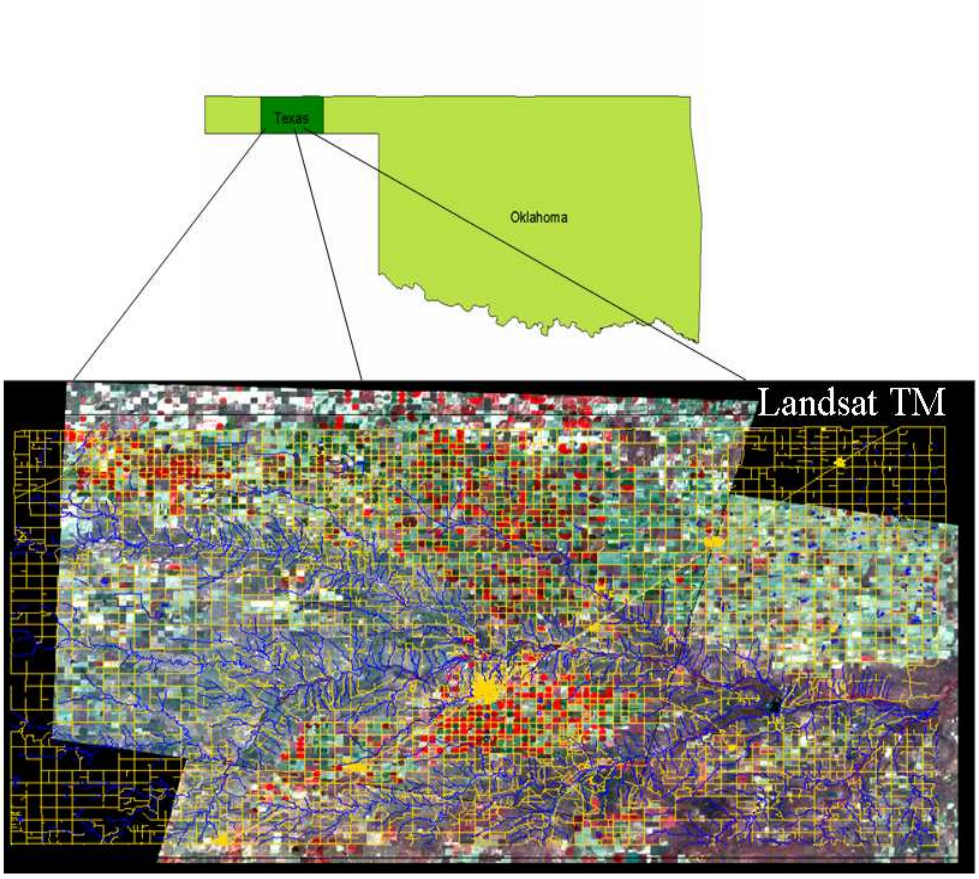


Figure 1.2: Texas County Landsat data superimposed with Road and Stream network information (Courtesy of Dr. Mahesh Rao of the Oklahoma State University's Geography Department).

a different resolution and is not useful in vegetation studies) were excluded. The following 20 layers are texture information that includes the local mean and local variance within a 3×3 window of each band in each season. The last 8 layers consist of different derived features like, Normalized Difference Vegetation Index (NDVI), Band Ratios and Band Differences. These are used for all the simulations conducted in the Chapters 3 and 4.

1.5 Organization and Contributions

Figure 1.6 gives a graphical representation of the succeeding chapters. Chapter 2 briefly describes the major algorithms used in this thesis. It describes the

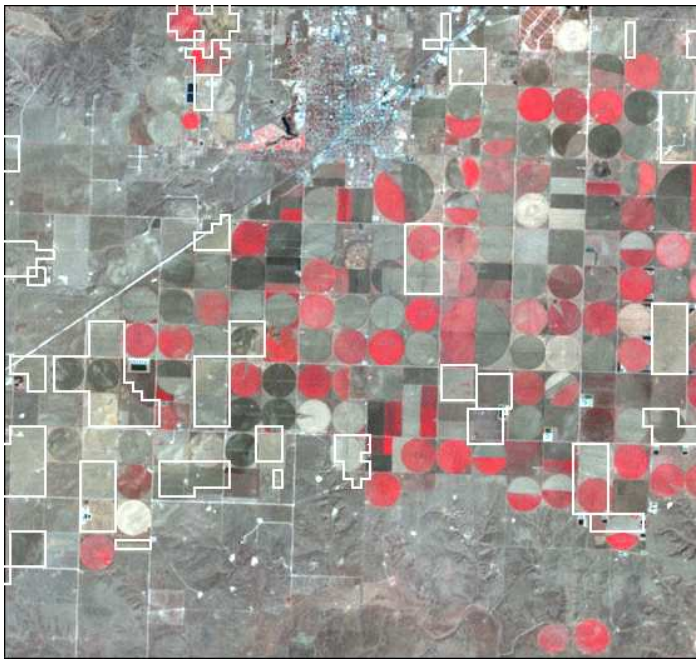


Figure 1.3: Clip of February 2000 Landsat TM image with superimposed CRP ground data(in white polygons).

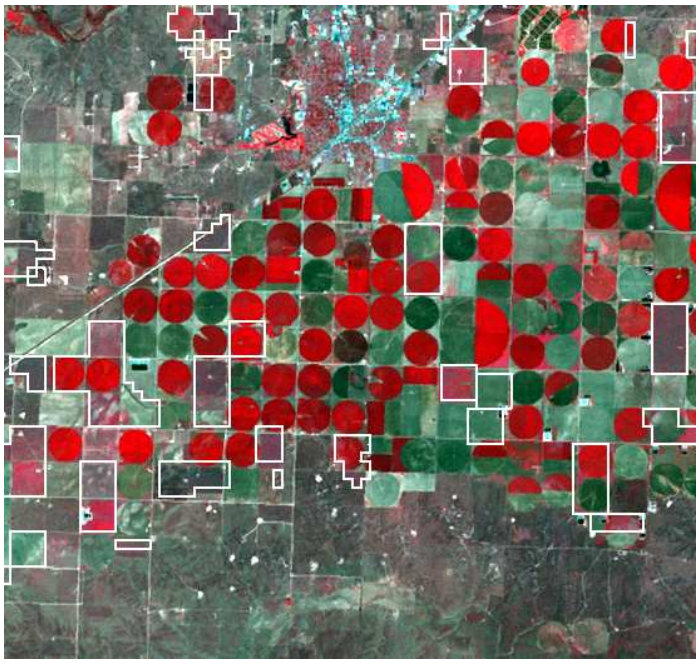


Figure 1.4: Clip of June 2000 Landsat TM image with superimposed CRP ground data(in white polygons).

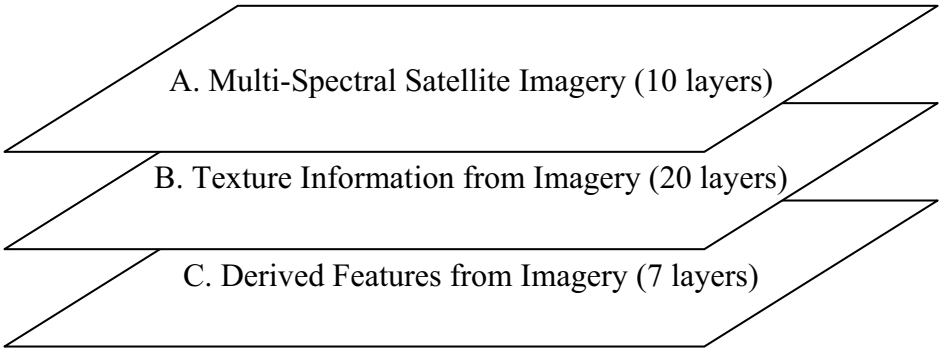


Figure 1.5: Different layers in each pattern.

SVM and the OCSVM. Also given are some details on the effects of kernel selection. Chapter 3 describes in detail the two methods discussed for compliance monitoring which are the distance based method and the ν insensitive method. Chapter 4 discusses the method for CRP mapping based on combinations of multiple OCSVMs. Chapter 5 is the conclusion.

The contributions of this thesis are mainly development of two methods for CRP compliance monitoring (discussed in Sections 3.4 and 3.5) and a new CRP mapping procedure (discussed in Chapter 4). Compliance monitoring methods are made by combining together the OCSVMs as the first stage and the general SVMs as the second stage. Difference lies in the method to select reliable training samples from the first stage for training the second stage. Two kernel space based methods are used for reliable sample estimation, one of which termed the *ν-insensitive approach* is proposed in this thesis. For CRP mapping we implemented a method using combinations of multiple OCSVM's trained on different CRP cover types. Also a pre-clustering procedure has been discussed for combining different CRP cover types prior to training.

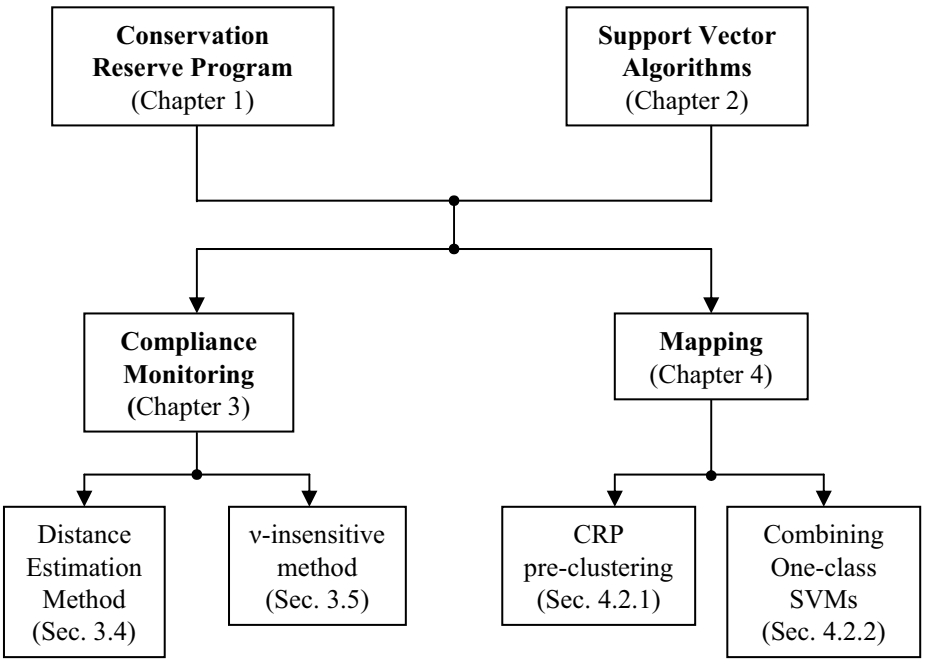


Figure 1.6: Thesis Organization.

Chapter 2

SUPPORT VECTOR ALGORITHMS

Support Vector Machines (SVMs) began to be widely used in the late 1990's. It is a learning algorithm that is based on the principle of margin maximization. SVMs have been widely applied in the fields of pattern recognition, regression analysis and density estimation, however we will be focussing on pattern classification as our problems fall in that domain. Comparative studies have found SVMs to perform as well as or better than most prevalent learning methods.

This chapter provides a brief review of general SVM; which is a supervised binary classifier; as well as the One-Class SVM (OCSVM) which is an unsupervised form of the SVM. Basic concepts of Statistical Learning Theory used for developing SVM's are discussed. Finally, we discuss kernel selection.

2.1 Statistical Learning Theory

The statistical learning theory forms the mathematical foundations for development of the Support Vector Algorithms. It deals with methods to quantify the risk in learning. It provides bounds on risk and an effective method to control the trade-off between risk and complexity of the classification function. This section is a brief review of concepts necessary to appreciate the Support Vector learning algorithm contained in [38], [5], [28] and [25].

2.1.1 General Learning Algorithm

The simplest problem in pattern recognition is the case where there are two classes, identified by class labels +1 and -1 respectively, requiring a supervised classification. The general SVM is a supervised binary classifier. We have a set of example data samples or what will be henceforth referred to as feature vectors x_i and corresponding labels y_i which form an independent and identical distribution related according to an unknown probability distribution $P(x, y)$. The training patterns are,

$$(x_1, y_1), ..., (x_l, y_l) \in R^N, y \in \{\pm 1\}, \tag{2.1}$$

where l is the number of patterns.

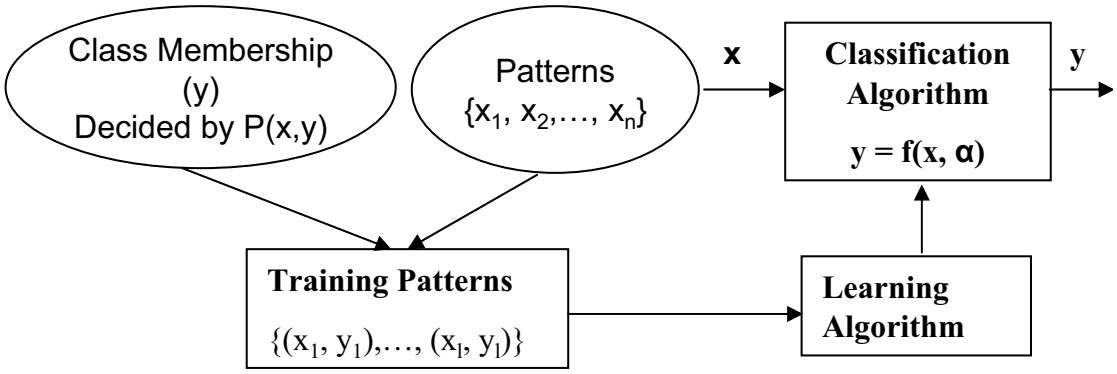


Figure 2.1: The general supervised binary($y \in \{\pm 1\}$) classifier. Our objective is to find the classification function $f(x, \alpha)$.

2.1.2 Risk and Risk bounds

Mapping $x_i \rightarrow y_i$ has to be learned. This mapping is defined by $f(x, \alpha) = y$, where $\alpha \in \Lambda$ is the set of function parameters also referred as the complexity of the function. As the number of function parameters increase, its complexity also increases. We want to learn the particular choice of α giving us the minimum risk. The expected risk is defined for a particular α as,

$$R(\alpha) = \int \frac{1}{2}|y - f(x, \alpha)| dP(x, y), \tag{2.2}$$

which cannot be calculated as $P(x, y)$ is unknown. Therefore for minimizing the risk the *empirical risk* is used which is defined as,

$$R_{emp}(\alpha) = \frac{1}{2l} \sum_{i=1}^{l} |y - f(x, \alpha)|. \tag{2.3}$$

The *expected risk* is what we get after classification. But during the training stage what we actually know is the *empirical risk*. It is possible that when $l \to \infty$, the *empirical risk* will converge to the *expected risk*. However for small sample sizes large deviations are possible and overfitting may occur. Therefore a small training error does not guarantee a small classification error.

Overfitting can be avoided by restricting the complexity of the function class $f(x, \alpha)$. The way for controlling function complexity is given in [38]. For the above defined learning problem for any $\alpha \in \Lambda$, $l > h$(h is defined in Section 2.1.3) and some η; so that $0 \leq \eta \leq 1$. Then the following bounds defined in [38] holds with a probability of at least $1 - \eta$;

$$R(\alpha) \leq R_{emp}(\alpha) + \phi(\frac{h}{l}, \frac{log(\eta)}{l}), \tag{2.4}$$

where the *confidence term* ϕ is defined as,

$$\phi(\frac{h}{l}, \frac{log(\eta)}{l}) = \sqrt{\frac{h(log(\frac{2l}{h}) + 1) - log(\frac{\eta}{4})}{l}}. \tag{2.5}$$

2.1.3 VC Dimension

The parameter h is called the VC (Vapnik-Chervonenkis) Dimension of a set of functions. VC dimension is a measure of the capacity of the function. Consider the case of two-class pattern recognition where the set of functions $f(x, \alpha) \in \{\pm 1\}$, for such a set of functions $f(\alpha)$ h is the maximum number of training points for which labels can be correctly assigned in all 2^l ways. Usually if the VC Dimension is h, at least one set of h points can be correctly assigned their labels, but it will not in general be true that all sets of points are correctly assigned. Thus the VC Dimension depends on the set of functions $f(x, \alpha), \alpha \in \Lambda$.

2.1.4 Structural Risk Minimization

From Equation 2.4 it can be seen that for a fixed number of training samples l, the risk is controlled by $R_{emp}(\alpha)$ and h. To control h a structure of nested subsets $S_n := f(x, \alpha) : \alpha \in \Lambda_n$ of $f(x, \alpha) : \alpha \in \Lambda$ is created such that;

$$S_1 \subset S_2 \subset .... \subset S_n \subset ..., \tag{2.6}$$

whose VC Dimensions as a result satisfy,

$$h_1 \leq h_2 \leq .... \leq h_n \leq .... \tag{2.7}$$

For a given set of observations as in Equation 2.1 the *Structural Risk Minimization principle* chooses the particular function subset S_n for which the guaranteed risk bound as in Equation 2.4 is minimal.

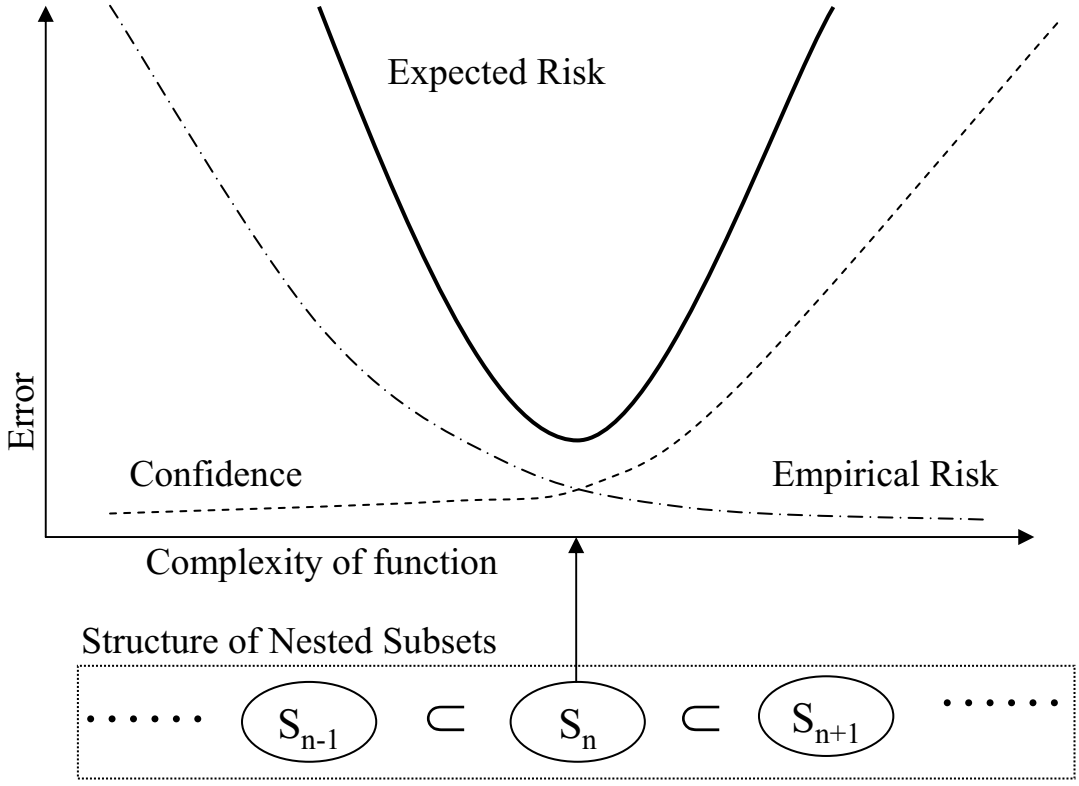


Figure 2.2: Shows the relation between Equation 2.4 and the sets determining function complexity (Equation 2.6). Thus illustrating *Structural Risk Minimization Principle*.

Figure 2.2 shows that, as the complexity of the function increases the empirical risk decreases and the VC Dimension increases. But the Expected Risk minimizes only for a certain particular subset S_n of the classification function, which if used will guarantee a good classification.

2.2 Support Vector Machine (SVM)

The SVM is a general purpose binary classification algorithm based on the margin maximizing principle. Reliance on the *Structural Risk Minimization principle* provides the theoretical background to guarantee a good classification accuracy. Here a brief review of basic SVM theory discussed in [38], [5], [28] and [25] is provided.

2.2.1 Linear Separable Case

This is the simplest case of pattern recognition which is easy to begin the explanation of the SVM algorithm. Here the feature vectors are assumed to be separable and can be separated by a linear decision boundary. The decision boundary is a hyperplane as the input space can be of any dimension. A hyperplane is a plane with respect to a feature space as it has one dimension less that feature space.

Now for the data set defined in Equation 2.1 it is possible that there will be a set of hyperplanes called *canonical hyperplanes* capable of separating the two classes of data. Our objective is to select the hyperplane separating the two classes of data with the maximum margin. The hyperplanes are of the form,

$$x \in R^N : (w.x) + b = 0, \tag{2.8}$$

for the above equation b is the offset form the origin and w is the normal to the hyperplane. Therefore the condition for classifying the data samples without error is

$$y_i((w.x_i) + b) \geq 1, i = 1, ..., l. \tag{2.9}$$

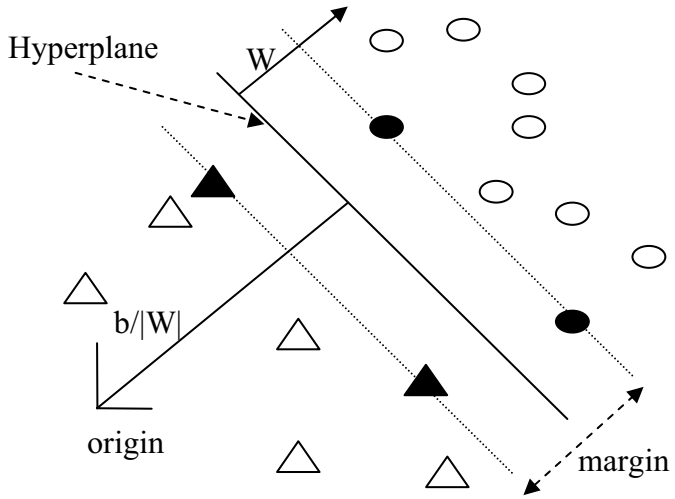


Figure 2.3: Diagrammatic representation of the linear separable case. Two dimensional case is shown for illustration purposes. Feature Vectors of the first class are represented as triangles and the other as circles. The hyperplane is in the middle of the margin between the two classes. The patterns lying on the margins (given in black) are the *Support Vectors* and are the ones that will be used in classification.

Consider two different patterns lying on the upper and lower margins (*Support Vectors*) i.e $(w.x_1) + b = 1$ and $(w.x_2) + b = -1$. The margin is therefore the distance between these two points measured perpendicular to the hyperplane i.e $(\frac{w}{||w||}.(x_1 - x_2)) = \frac{2}{||w||}$ thus the best hyperplane can be found by maximizing this margin or by minimizing,

$$\tau(w) = \frac{1}{2}||w||^2, \tag{2.10}$$
$$\text{subject to: } y_i((w.x_i) + b) \geq 1, i = 1, ..., l.$$

2.2.2 Relation with VC Dimension

As defined in [38] for a class of hyperplanes, for example of the form $f(x) = sgn((w.x) + b)$ the VC-Dimension h can be bounded in terms of;

$$h \leq R^2\Lambda^2 + 1, \tag{2.11}$$

where Λ is an upper bound constraining the length of the weight vector of the hyperplane in canonical form, and R is the radius of the smallest sphere containing

the data in the space where the hyperplane is constructed. The smaller this sphere, the smaller is the capacity. Thus it can be said that by requiring a large lower bound on the margin (i.e. a small Λ) we obtain a small VC-dimension. By allowing for separations with small margin we can potentially separate a much larger class of problems.

2.2.3 Linear Non-separable Case

In most classification cases data are not separable. In terms of the SVM classification it means that the margin region cannot be free of feature vectors. The hyperplanes accommodating this property are referred to as *Soft Margin Hyperplanes*. To make this possible slack variables have to be introduced,

$$\xi_i \geq 0, i = 1, ..., l. \tag{2.12}$$

The equation (2.9) is transformed to,

$$\begin{aligned} \tau(w, \xi) &= \frac{1}{2}||w||^2 + C\sum_{i=1}^{l}\xi_i, \\ \text{Subject to: } & y_i((w.x_i) + b) \geq 1 - \xi_i, i = 1, ..., l. \end{aligned} \tag{2.13}$$

Literally this approach allows some misclassification of the data to obtain a linear classification boundary. The parameter C called the *regularization constant* determines the penalty on the errors. The second term in the Equation 2.18 controls the degree to which the misclassified data will affect the formation of the decision boundary.

2.2.4 Optimization

To solve the convex optimization problem defined in Equations 2.10 and 2.9 a langragian L is formed. The Equation 2.10 now becomes;

$$L(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{l}\alpha_i(y_i((x_i.w) + b) - 1), \tag{2.14}$$

with langragian multipliers $\alpha_i \geq 0$. The langragian L has to be maximized with respect to α_i and minimized with respect to w and b. The condition that at the saddle point the derivatives of L with respect to the primal variables (w and b) must vanish,

$$\frac{\partial}{\partial b} L(w, b, \alpha) = 0, \frac{\partial}{\partial w} L(w, b, \alpha), \tag{2.15}$$

leads to,

$$\sum_{i=1}^{l} \alpha_i y_i = 0 \text{ and } w = \sum_{i=1}^{l} \alpha_i y_i x_i. \tag{2.16}$$

According to Kuhn-Tucker theorem from optimization theory, at the saddle point α_i can be nonzero only for points x_i which satisfy,

$$\alpha_i [y_i((x_i.w) + b) - 1] = 0, i = 1, ..., l. \tag{2.17}$$

These points are called *Support Vectors* and in the Figure 2.3 lie exactly on the margin. These are the only feature vectors required for classification.

Substituting Equation 2.16 into 2.14 the dual form (used for solving for the support vectors) of the optimization is derived:

$$\max W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j (x_i.x_j) \tag{2.18}$$

$$\text{Subject to: } \alpha_i \geq 0;\ i = 1, ..., l;\ \sum_{i=1}^{l} \alpha_i y_i = 0.$$

On substituting the above, the decision function is obtained as,

$$f(x) = sgn(\sum_{i=1}^{l} \alpha_i y_i (x, x_i) + b). \tag{2.19}$$

2.2.5 Non-Linear Support Vector Machines

Non-Linear separation boundaries are needed for solving general purpose problems. SVM tries to solve this problem by non-linearly transforming the input feature space by a mapping function $\Phi : x_i \rightarrowtail z_i$ into a high dimensional feature

space where a linear separation is done. This is shown in Figure 2.4. A decision function of the following form is obtained;

$$f(x) = sgn(\sum_{i=1}^{l} \alpha_i y_i (\Phi(x).\Phi(x_i)) + b), \tag{2.20}$$

where $(\Phi(x).\Phi(x_i))$ are dot products computed in the projected space.

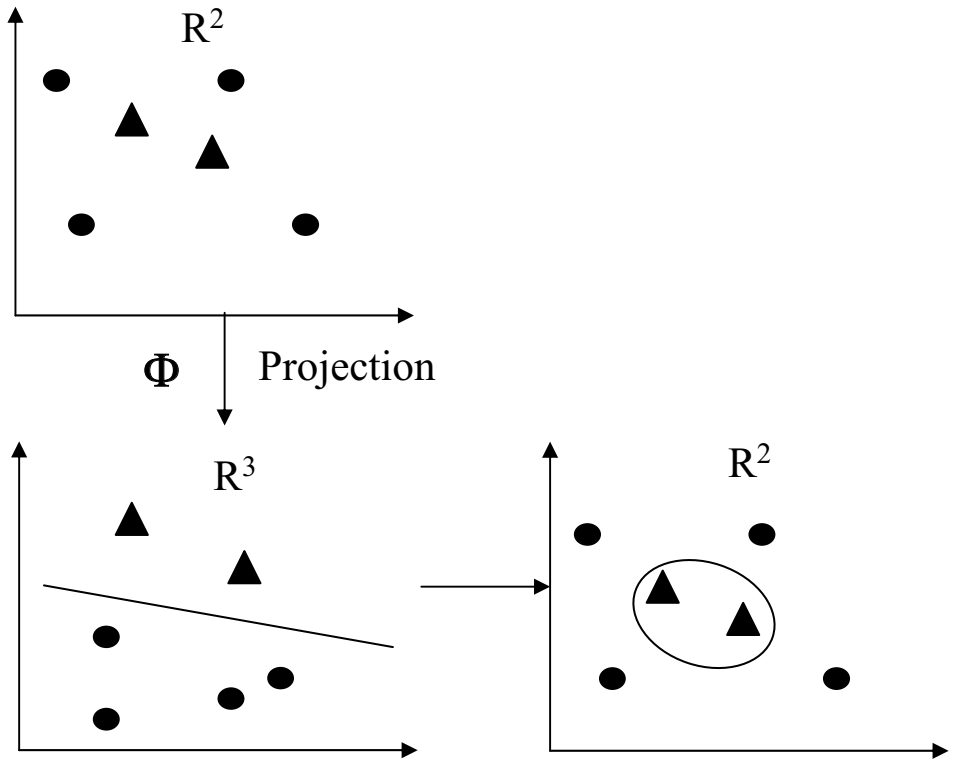


Figure 2.4: Non-Linear decision boundaries are constructed by projecting the data to a higher dimensional space, where a linear classification boundary is constructed.

2.2.6 Support Vector Classification

From the training procedure the Support Vectors(SV's) are known. For each input vector x the kernel distance is calculated and based on the decision function the vector is assigned to either of the classes. Figure 2.5 shows the case where there are three SV's.

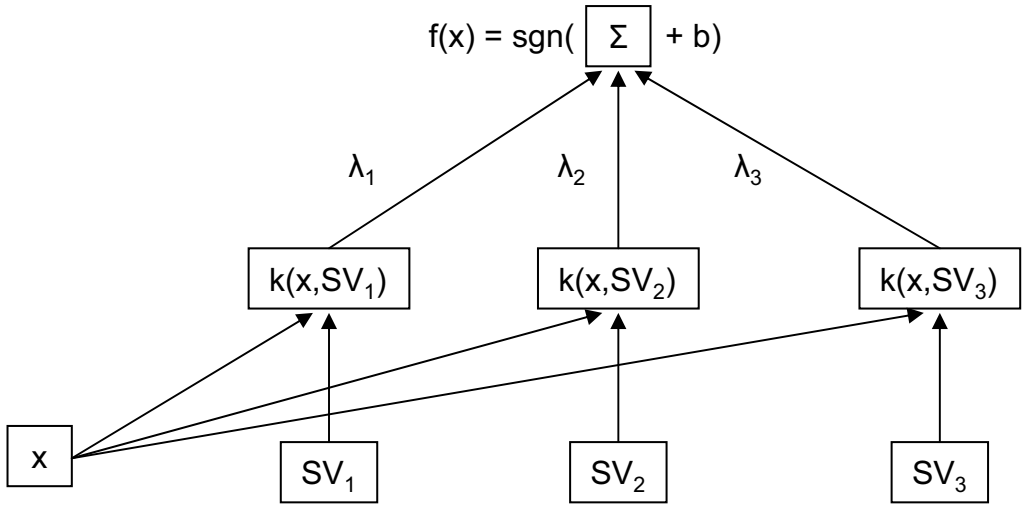


Figure 2.5: Here x is an input vector. There are three *Support Vectors*. The kernel function is evaluated with respect to each of them. The classification function (Equation 2.20) decides to which of the classes ± 1 the input x belongs.

2.3 One-Class Support Vector Machine(OCSVM)

This is an extension of the regular SVM to the case of unsupervised classification. Here the labels of the training data sample are unknown. In this section two different approaches to this problem are discussed.

2.3.1 One-Class Classification

In one-class classification only the information about the target class is available. The boundary between the data of the target class, which is provided, and all other data, considered as outliers has to be estimated based on the data of the target class alone. This method is also referred to as novelty detection.

One-Class classification will be introduced with the example given in Figure 2.6 described in [34]. Here training patterns of data about apples and pears are available. Each object has two feature values viz. weight and width. The two classes can be separated without errors by the solid line in Figure 2.6, which is the

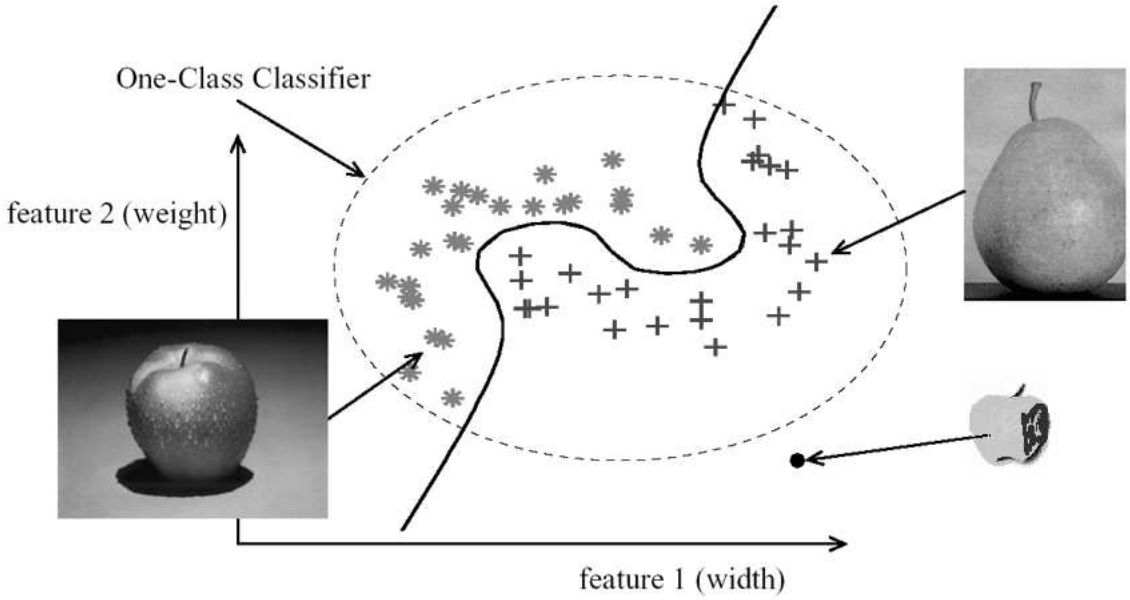


Figure 2.6: Figure from [34] shows the difference between the regular classifier and the One-class classifier.

normal two-class classifier. Consider now that a new pattern of a rotten apple in the lower right corner is introduced. It cannot be distinguished from the pears. Thus for a two-class classifier all patterns have to be either regular apples or pears and anything else won't be classified correctly. However the one-class classifier (denoted by dotted line) can differentiate the outlier after training.

2.3.2 Hyperplane based model

This algorithm for OCSVM has been proposed in [30]. It tries to estimate the region in the projected feature space where majority of the data resides.

Here the objective is to separate the data in the kernel space from the origin with the maximum margin. So it amounts to finding some hyperplane $w \in F$ that separates the unlabelled training data from the origin with the threshold ρ. The function $f_w(x) = (w.\Phi(x))$ has to be determined and the pattern x belongs to *one class* when $f_w(x) \geq \rho$. To separate the data set from the origin, the following

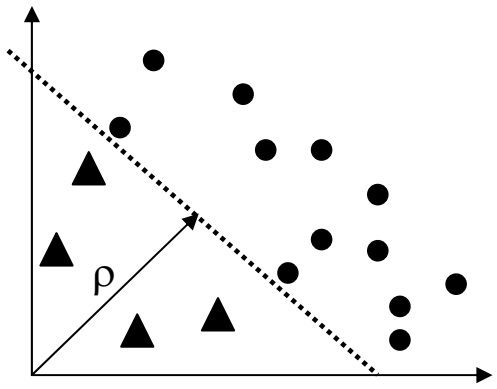


Figure 2.7: Triangular objects are outliers and circular objects show the majority data. Hyperplane is given by the dotted line.

quadratic program has to be solved,

$$\min_{w\in F,\xi\in R^l,\rho\in R} \frac{1}{2}||w||^2 + \frac{1}{\nu l}\sum_{i=1}^{l}\xi_i - \rho, \tag{2.21}$$
$$\text{subject to:}(w.\Phi(x)) \geq \rho - \xi_i \ , \ \xi_i \geq 0.$$

This decision function becomes of the form,

$$f(x) = sgn((w.\Phi(x)) - \rho), \tag{2.22}$$

which will be positive for most examples in the training set. The parameter $\nu \in (0, 1)$ controls the percentage of outliers.

The dual form for this optimization is obtained as,

$$\min_{\alpha} \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j k(x_i, x_j), \tag{2.23}$$
$$\text{subject to } 0 \leq \alpha_j \leq \frac{1}{\nu l}, \sum_i \alpha_i = 1.$$

At the optimum, the two inequality constraints in Equation 2.21 become equalities if $0 \leq \alpha_j \leq \frac{1}{\nu l}$. Therefore ρ can be recovered by exploiting that for any such α_j the corresponding pattern x_i satisfies,

$$\rho = (w.\Phi(x_j)) = \sum_j \alpha_j k(x_j, x_i). \tag{2.24}$$

In this algorithm ν is defined as an upper bound on the fraction of outliers and a lower bound on the fraction of SV's. For practical implementations it can be roughly taken as the percentage of outliers.

2.3.3 Hypersphere based model

This algorithm for OCSVM has been proposed in [34] and [35]. It is also referred to as *Support Vector Data Description*. Here the objective is to create a hypersphere with the minimum volume in kernel space containing all the data. The sphere is defined by a center a and a radius $R > 0$. The volume is minimized by minimizing R^2 and the sphere has to contain all the training objects x_i.

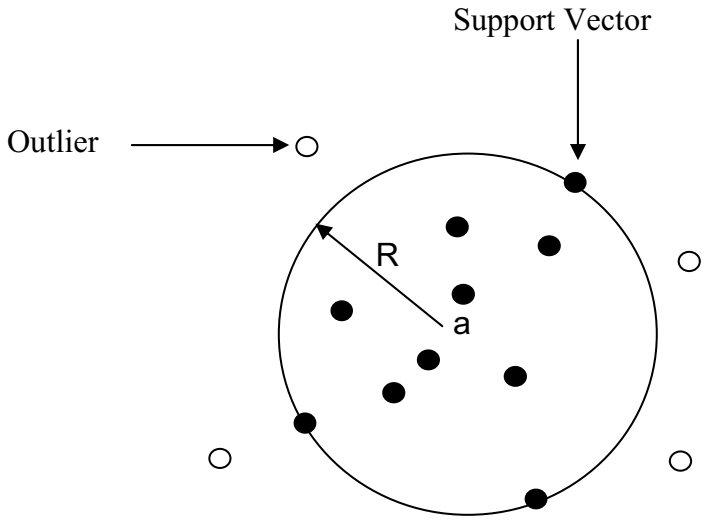


Figure 2.8: White objects are outliers and black objects show the majority data. R is radius and a is the center of the hypersphere.

The error function to minimize is:

$$F(R, a) = R^2, \tag{2.25}$$

$$\text{subject to: } ||x_i - a||^2 \leq R^2, i = 1, ..., l. \tag{2.26}$$

To allow for the possibility of outliers in the training data, slack variables $\xi_i \geq 0$ are introduced. The distance from x_i to the center a should be strictly smaller than R^2, but larger distances will be penalized. So the minimization problem changes to,

$$F(R, a) = R^2 + C \sum_l \xi_i. \tag{2.27}$$

$$\text{subject to: } ||x_i - a||^2 \leq R^2 + \xi_i, \xi_i \geq 0, \forall i. \tag{2.28}$$

where the parameter C controls the tradeoff between the volume and the errors. C is similar to the ν of the hyperplane model.

Utilizing Lagrangian multipliers $\alpha_i \geq 0$ and setting the derivatives to 0, the value of R can be obtained as,

$$R^2 = (x_k.x_k) - 2\sum_l \alpha_i(x_i.x_k) + \sum_{i,j} \alpha_i\alpha_j(x_i.x_j), x_k \in SV, \tag{2.29}$$

where SV is the set of support vectors (feature vectors lying at the boundaries of the sphere) which have $\alpha_i > 0$.

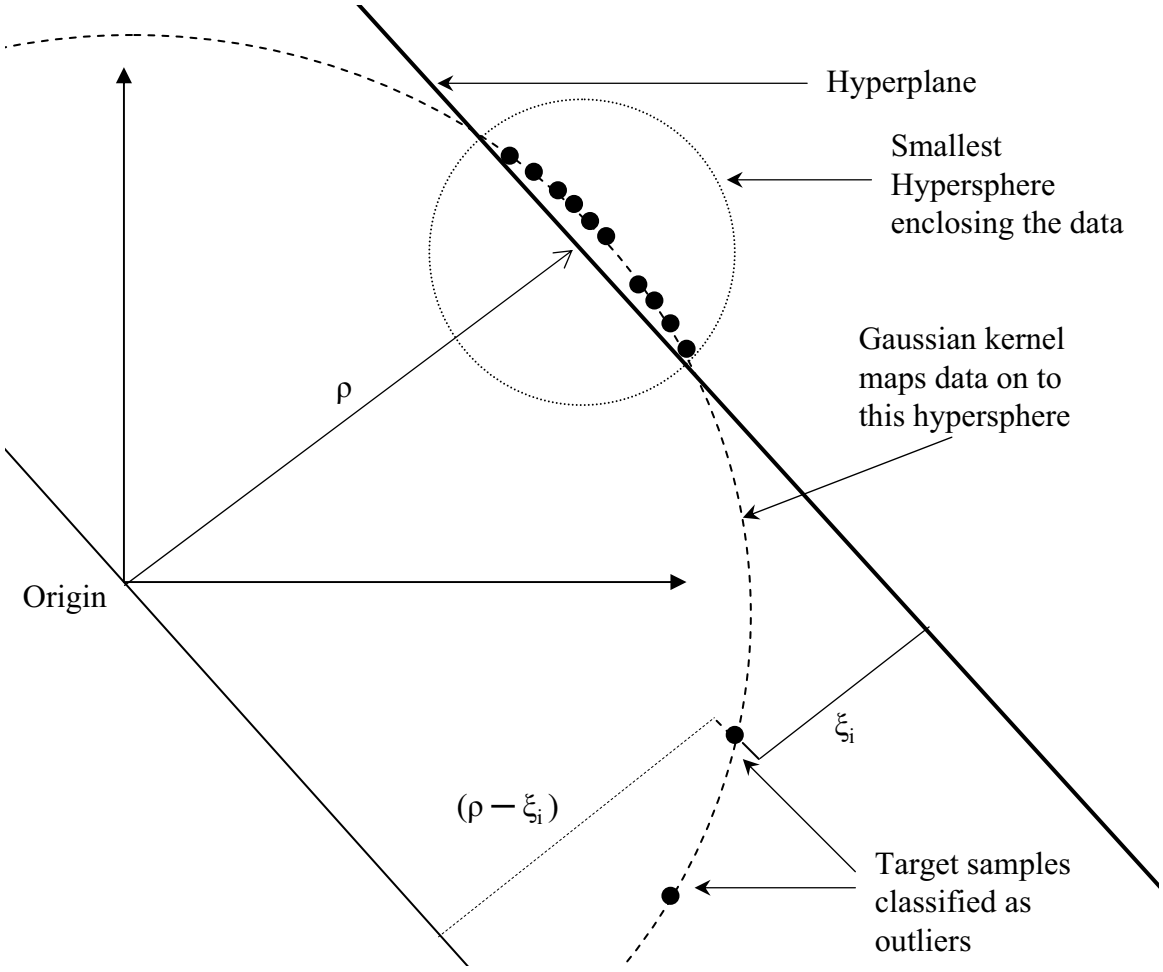


Figure 2.9: Equivalance between the hypersphere and hyperplane model for OCSVM from [37].

2.3.4 Model Equivalence

Using the Gaussian kernel, data is preprocessed to have a unit norm. Thus feature vectors lie on the surface of a unit hypersphere in kernel space. In [30], [34] and [35] both methods have been proved to have comparable solutions when the

Gaussian kernel is used. A diagrammatic representation is shown in Figure 2.9. For non-normalized data the solutions become incomparable due to the differences in model description. Among all the commonly used kernels the Gaussian kernel gives tightest boundary descriptions and hence the best classification results as described in [34].

2.4 Kernel Selection

The kernel is effectively a mapping function that does the transformation $\Phi : x_i \rightarrowtail z_i$ into a higher dimensional space. The projected space H is a very high dimensional one or in some cases an infinite dimensional space. It can be seen that maximizing the target function and evaluating the decision functions require the computation of dot products $(\Phi(x).\Phi(x_i))$. Under Mercer's conditions these expensive computations can be reduced significantly by using a suitable function k such that;

$$(\Phi(x).\Phi(x_i)) = k(x.x_i), \tag{2.30}$$

where k is the kernel. As only dot products are required for the solution, it can be obtained using kernels without even knowing the mapping Φ. Now the decision function (2.20) is converted to,

$$f(x) = sgn(\sum_{i=1}^{l} \alpha_i y_i.k(x.x_i) + b). \tag{2.31}$$

The *curse of dimensionality* in statistics says essentially that the difficulty of an estimation problem increases drastically with the dimension N of the space. Here we are not dealing with actual data projections into a high dimensional feature space. *Statistical Learning Theory* [38] tells that the contrary can be true; it is not the dimensionality of the feature space but the complexity of the classification function that matters.

The generally used kernels are;

- Polynomial Kernel :$k(x, x_i) = (x.x_i)^d$,
- Gaussian Kernel :$k(x, x_i) = exp(-\gamma||x - x_i||^2)$,
- Sigmoid Kernel :$k(x, x_i) = tanh(m.(x.x_i) + \Theta)$.

2.4.1 Feature Spaces Induced by Kernels

The important thing to understand is data is not physically projected into the high dimensional space. Only the optimization of the SV algorithm is done as if the data were in high dimensional space. As the optimization consists of dot products it is possible to use kernels to simplify the computation.

The kind of kernels used in SV algorithms are Mercer Kernels. The space to which data ia projected using kernels is described as a *Reproducing Kernel Hilbert Space*. Detailed description of the kernel space properties have been given in [29].

The kernel that is normally used is the Gaussian Kernel. This kernel allows tighter decision boundaries and thus provides better classification. In [34] it was found to be the best kernel for both types of OCSVMs. This kernel has a width parameter γ which has to be tuned by the user. This kernel is independent of the position of the patterns with respect to the origin, it only utilizes the distances between the patterns. Also the influence of the norms is avoided. Patterns are mapped to unit norm vectors (norm of mapped objects, $\Phi(x_i).\Phi(x_i) = 1$), only the angles between the patterns count.

2.4.2 Study on Gaussian Kernel

As the Gaussian Kernel is the most widely used a lot of study has been done to understand its working and behavior. [20] provides great insight into how the kernel parameters affect the classification.

When $\gamma \longrightarrow 0$ the Gaussian kernel can be represented as,

$$\begin{aligned} K(x_i, x_j) &= exp(-\gamma||x_i - x_j||^2) \\ &= 1 - \gamma||x_i||^2 - \gamma||x_j||^2 + 2\gamma x_i^T x_j + o(2\gamma||x_i - x_j||^2). \end{aligned} \quad (2.32)$$

From the above equation it can be seen that the kernel distances between the different patterns will be close to 1. So the patterns in the projected space will be more spread out than otherwise. SVM's trained using this kernel will cause underfitting. Thus less Support Vector's will be needed to classify the data.

It can be see that in the opposite case of when $\gamma \longrightarrow \infty$ the kernel distances are smaller, and the patterns are more clustered together in the kernel space. In this case overfitting will occur and a much larger number of Support Vector's will be needed to classify the data.

So no fixed γ value will fit all problems. Thus γ values have to be adjusted on a problem specific basis and the definition of accuracy for the specific problem. The usual approach is cross-validation. Here the available training samples are divided into sets and are trained separately using different γ values and the one giving the best training result is selected as the appropriate value.

2.5 Applications in Remote Sensing

Our problem comes under the domain of Landcover Classification using Landsat TM images. SVM is used for this problem domain in [16], [21], [27], [33] [17] and [4]. In [16] a joint algorithm combining the two-class SVM and contextual classification using the Iterated Conditional Modes algorithm was proposed. This joint classifier was found to perform better compared to maximum likelihood, Gaussian mixture models, 1-nearest neighbor and an SVM. Comparisons of different supervised algorithms ware performed in [21], [27], [33] and [17]. The algorithms tested were the SVM, iterated conditional modes, maximum likelihood classifier, Decision Tree Classifier and neural networks classifier. The algorithms were tested on different problems of land cover classification using Landsat TM images. It was concluded

that the SVM provided the most statistically accurate results for each of their test cases. However the SVM results appear noisy as it is not a region based clustering algorithm. Also the SVM was found to give the best performance when low amounts of training data were used. Usage of the SVM algorithm to solve linear spectral mixture models were checked in [4]. It is argued that the SVM framework based on margin maximization is more appropriate for empirical mixture modeling, as nonseparable distributions of pure classes can be handled appropriately, as well as nonlinear mixture modeling.

Chapter 3

CRP COMPLIANCE MONITORING

3.1 Introduction

USDA is faced with the problem of farmers not maintaining CRP tracts according to contract stipulations. Current methods for CRP compliance monitoring involve intensive manual inspection of aerial photographs which is time-consuming and costly. USDAs Common Land Unit (CLU) data used for general compliance issues is generated from aerial photographs with a resolution about $1m \times 1m$, which are updated every 1-2 years and may not be very efficient for CRP compliance monitoring on a large scale . In addition, existing CRP reference data obtained from USDAs Natural Resource Conservation Service (NRCS) is not very accurate or up-to-date for the management purpose. There is a need of an automatic compliance monitoring method which can examine CRP tracts more efficiently and promptly with minimum human involvement. Two methods are discussed for this problem.

3.2 Problem Definition

We are tackling our problem by dealing with each CRP tract individually. We include some non-CRP areas around the boundaries of each CRP tract for reference purposes. We assume that the majority of region ($> 50\%$) within a CRP tract is compliant, which is true for most CRP tracts. Each CRP tract can now be thought of as single large class of data covering majority area belonging to the CRP class and a mixture of various smaller classes of non-CRP areas (outliers).

Constrained by these conditions our problem reduces to one of finding the percentage of the majority cluster lying mostly within the area said to be having CRP cover type. Based on the CRP reference data, if the classification accuracy is high, then we think the CRP tract is compliant. While if the classification accuracy is low, there exists higher probability that the CRP tract is non-compliant.

3.3 General Strategy

Our solution involves incorporating both SVM and OCSVM. Due to the fast changes in CRP enrollment and expiration during the past years, the present CRP reference data is not accurate enough for management purpose. Consequently, CRP compliance monitoring based on the Landsat imagery has to be considered via an semi-supervised way, where the existing CRP reference data can only provide locality information to select suitable plots for checking for compliance.

OCSVM (Section 2.3) is first used to separate the majority data from minor outliers in a tract where the majority is assumed to be the CRP area. The OCSVM results are used to train a SVM to further refine the previous results. This is done because the OCSVM produces a complex decision boundary marked by a large number of Support Vectors, whereas the SVM provides a more natural decision boundary. Also the usage of data from two classes (CRP and non-CRP) improves the overall performance. The CRP reference data is used as the baseline of compliance issue where a high classification accuracy with respect to the reference data indicates the compliance of CRP enrollment, while a low accuracy implies the possibility of non-compliance.

The difference between both methods lies in choosing the training samples for training of the SVM as in step 3 of Figure 3.1. This is difficult because we do not know anything about the outliers like their samples, class membership and more importantly about their percentage within a specific clip chosen to check for compliance.

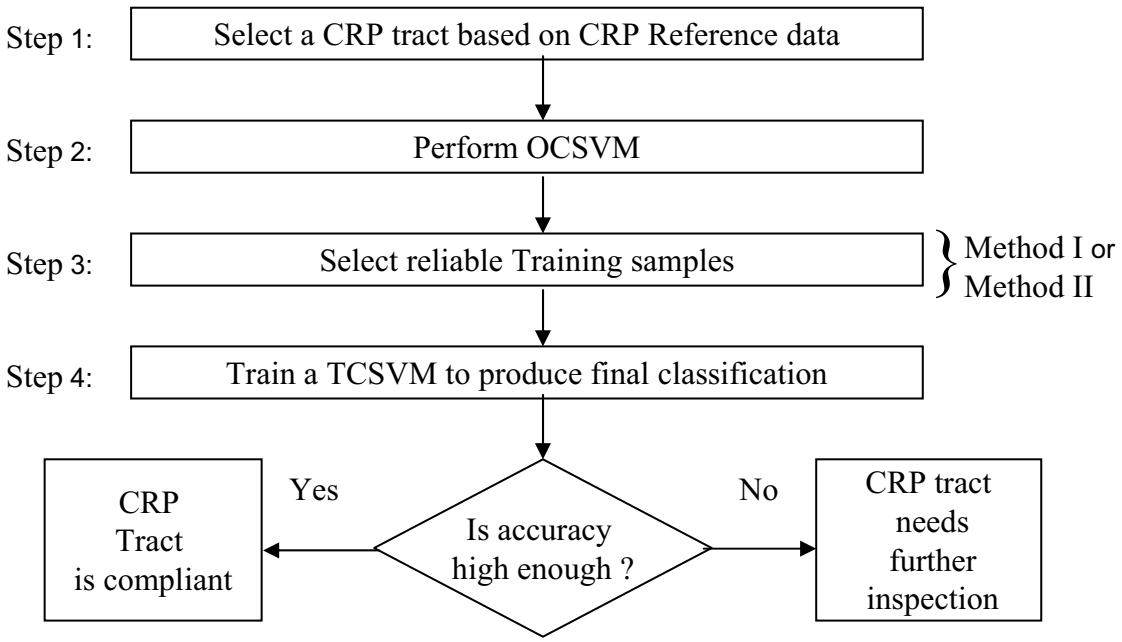


Figure 3.1: Flowchart of general strategy.

3.4 Method I [1]

3.4.1 Introduction

In this method the reliable training patterns are selected through a two step process. Initially the model selection is done for the OCSVM. This is done via a heuristic method discussed in [26] and discussed in Section 3.4.2. The spatial properties of the data are used to decide the more reliable patterns among the ones classified belonging to both the majority (CRP) and outlier or minority (non-CRP) class. This method has been discussed in [6].

3.4.2 Model Selection for OCSVM

Model selection aims at determining ν in OCSVM(Section 2.3.2). $\nu \in (0, 1)$ can be ideally set to the fraction of outliers which is unknown for our problem. A

[1] `Appeared in [6].`

simple and effective heuristic approach for estimating ν was proposed in [26]. This method works well in cases where the majority and the outliers are well separated.

The idea is to initially try out the classification for different ν values and then to select the one that has the largest separation distance between the classified majority and outlier class in kernel space. The separation distance for a particular value of ν is computed as;

$$D_\nu = \frac{1}{N_+} \sum_{f_w(x) \geq \rho} f_w(x) - \frac{1}{N_-} \sum_{f_w(x) < \rho} f_w(x) \tag{3.1}$$

where N_- and N_+ are the number of patterns in each class and $f_w(x) = (x.w) + b$, which calculates the distance from the origin. As defined in Equation 3.1 D_ν provides an average estimation of the distance between two patterns in the kernel space, and the specific ν value which provides the largest D_ν should have better separability among majority and outliers compared with other ν values.

3.4.3 Proposed Algorithm

OCSVM gives us an initial classification of CRP and non-CRP areas. To get a more natural decision boundary we sample the OCSVM results to get more reliable training samples for CRP and non-CRP regions. Then we train a SVM to reclassify the whole clip. The flowchart of this process is illustrated in Figure 3.2.

In this flowchart, the first step is to construct a clip based on the CRP reference data (Section (1.4)), where the majority of the clip belongs to a single CRP type with some surrounding non-CRP areas. In order to get the training samples for OCSVM, we uniformly sample the CRP tract with a 30% rate, which means 30% of CRP samples are used for training. Then the OCSVM is trained based on a given ν value, and Equation 3.1 is calculated to get the average inter-class distance. OCSVM is trained several times with different ν, and the one with the largest inter-class distance is selected. The trained OCSVM is applied to the whole clip to get a segmentation map. Afterward, reliable training samples for SVM

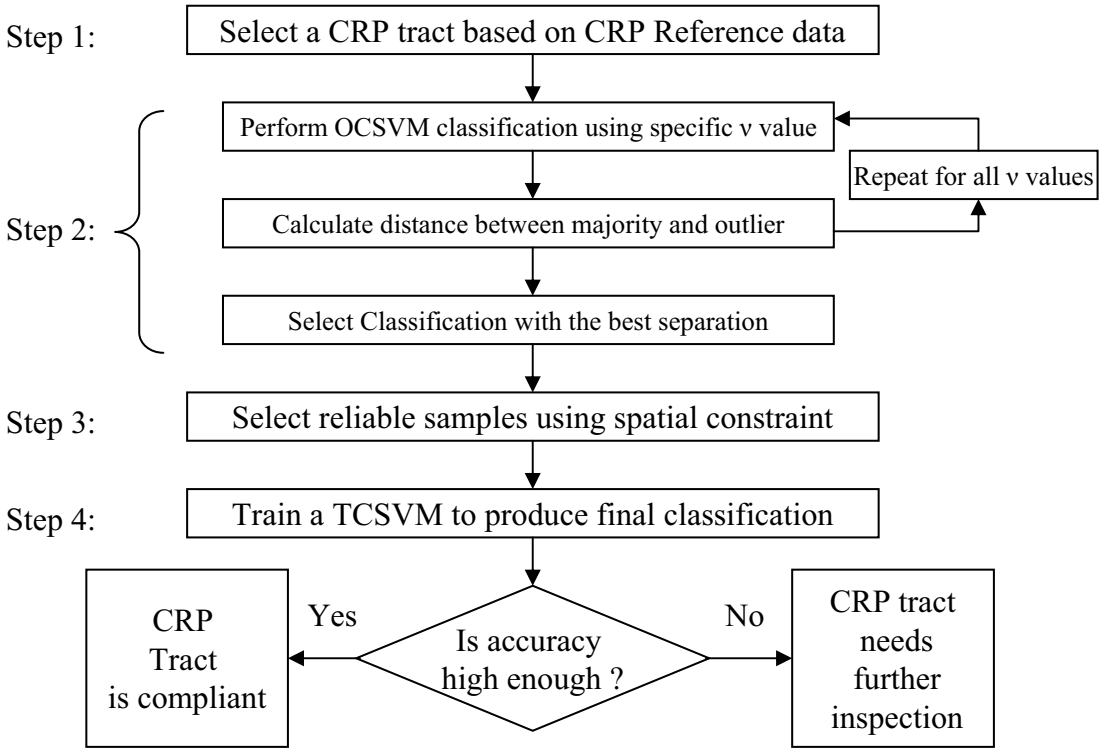


Figure 3.2: Flowchart of Method I.

are selected based on the OCSVM result. A sample is considered to be reliable if all samples within its 5×5 neighborhood have the same labels. The whole clip can be reclassified using the trained SVM. Finally, the accuracy of the final classification is used to determine if the CRP tract is in compliance with the enrollment stipulations.

3.4.4 Experimental Setup

A clip of the derived data set from Section (1.4) is chosen such that the majority of the feature vectors belong to a single CRP type as specified in the CRP reference data. OCSVM and SVM implementations found in [2] are used in our simulation.

For the OCSVM, the Gaussian kernel with $\gamma = 0.000001$ is used according to the cross validation results. The ν value varies from 0.05 to 0.5 in steps of 0.05 to estimate the best possible ν. More accurate estimation of ν can be obtained via a coarse to fine process with more computation time. The OCSVM classification

is performed based on the selected ν value. Then as mentioned in Section 3.4.3, OCSVM results are re-sampled for SVM training via a 5×5 window operation. The Gaussian kernel is also used for SVM and cross validation shows that $\gamma = 0.1$ is the preferred kernel width.

3.4.5 Simulations and Discussions

Simulations are performed on four different clips as shown in Figure 3.3. In the simulation, ν value is first determined for each clip based on the separation distance, and the results are shown in Table 3.1. Figure 3.3 (a) illustrates the February Landsat images of four clips. Figure 3.3 (b) illustrates the June Landsat images of four clips. (c) is the CRP reference data of these clips, where light grey areas are CRP regions and black areas are non-CRP areas. Figure 3.3 (d) shows the OCSVM results, and (e) indicates the area for SVM training, where light grey areas are the reliable ones that SVM training samples can be selected, and dark grey areas are those that are rejected as unreliable samples for SVM training. Figure 3.3 (f) is the classification results after using SVM. These four clips are selected such that around $\frac{1}{3}$ of the area are non-CRP samples.

Table 3.1: Kernel space inter-cluster separation distance at different ν values. Largest distances for each clip is in bold.

ν	**Clip − 1**	**Clip − 2**	**Clip − 3**	**Clip − 4**
0.05	0.00315	0.00614	0.00095	0.00142
0.1	0.00554	0.00771	0.00244	0.00255
0.15	0.00764	0.00841	0.00328	0.00372
0.2	0.00900	0.00613	0.00410	0.00503
0.25	0.00988	0.00856	0.00435	0.00630
0.3	0.01105	0.01087	0.00511	0.00663
0.35	0.01137	0.01174	0.00586	0.00822
0.4	0.01175	**0.01252**	0.00633	0.00922
0.45	0.01225	0.00882	**0.00730**	0.00989
0.5	**0.01312**	0.01106	0.00724	**0.01093**

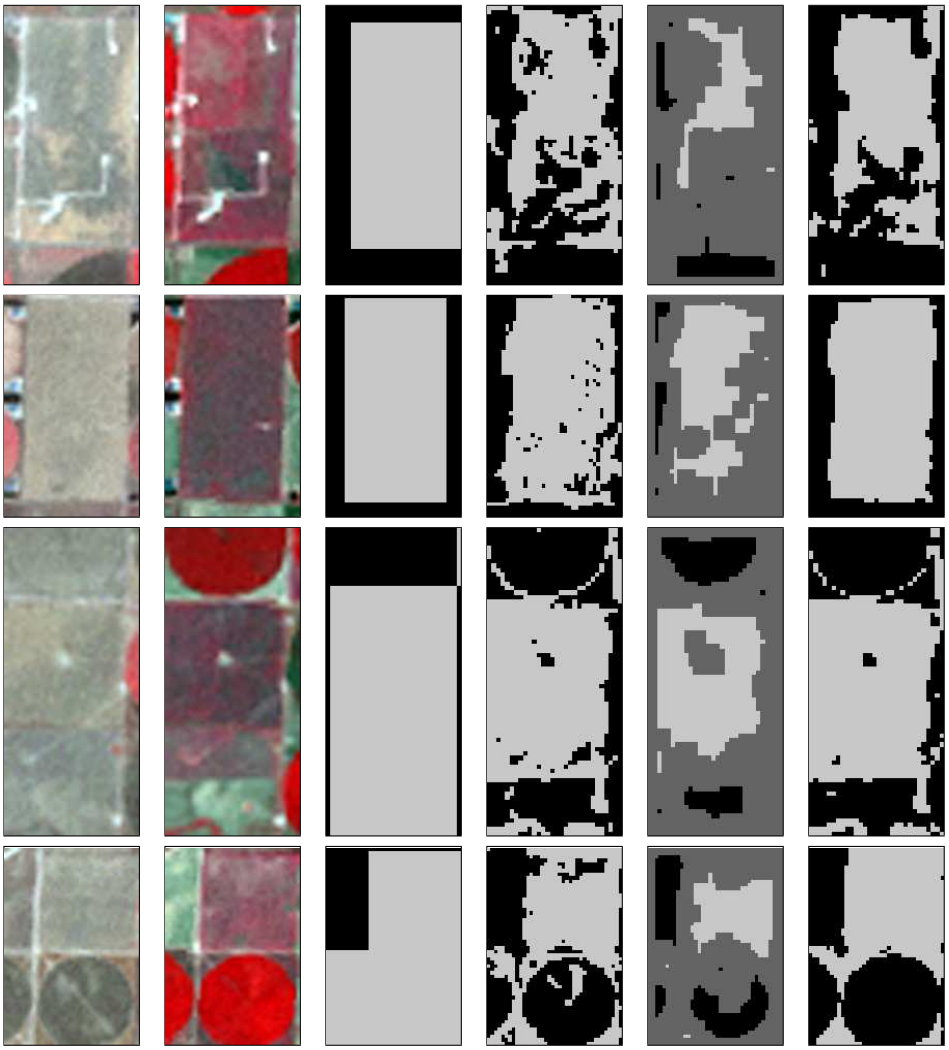


Figure 3.3: Simulation results. (a) Landsat Feb. 2002 images: Clip-1 to Clip-4 from top to bottom. (b) Landsat June 2002 images. (c) CRP reference data. (d) OCSVM results. (e) Resampling areas. (f) Final SVM classification results.

From the Figure 3.3 the improvements due to Spatial Processing and SVM classification is obvious by comparing the images in (d) and (f) groups. A good amount of noisy class labels have been reduced resulting in a more natural looking classification. Clip-2 from Figure 3.3 follows the assumption that the majority of the clip is the compliant CRP area with a single cover type, thus the classification result is very close to the reference data, resulting in a high accuracy (95.39%). It had an improvement of 7.3% due to spatial processing followed by SVM. Clip-1 has some build up area and some variations in the cover type so the accuracy is lower (78.39%) according to the reference data. There is an improvement of 4.2% due to spatial processing and application of SVM. Clip-3 has some CRP land bordered by some agricultural lands. There seems to be different cover types on the CRP land so the accuracy is even lower than Clip-1 (75.31%). Clip-4 is non-compliant as can be seen in the image (dark circular regions indicate agriculture being practiced with pivot irrigation systems). So here the classification accuracy is poor (59.25%). As can be seen when there is active agricultural land the classification accuracy becomes low implying a possible non-compliance problem.

3.5 Method II [2]

3.5.1 Introduction

Method I (Section 3.4) suffers when the two clusters are not clearly separable, which occurs quite often in large scale remotely sensed data. Here a ν- insensitive approach is presented where a mild deviation from true ν, which is unknown, will not significantly affect the classification performance. This method makes use of pattern distribution within the kernel space to decide on the reliable training patterns. ν-insensitivity is achieved by selecting sufficient and reliable training samples according to their position in kernel space (which does't change with ν values) with respect to the hyperplane obtained from the OCSVM. Compared with distance estimation method from Section 3.4.2, this method reduces the computational load by avoiding

[2] Joint work with Xiaomu Song. Presented in [32].

ν estimation, and also improves the robustness of the classification performance. By comparing the classification results with the CRP ground data, the compliance issue can be addressed. This method has been discussed in [32].

3.5.2 Study of Kernel Space

Since ν is the upper bound of the amount of outliers, changing ν actually changes the position and orientation of the classification hyperplane in the feature space. An improper ν would cause some outliers to be misclassified as the majority class, or vice versa. Patterns which are prone to be misclassified, are usually located around or on the optimal hyperplane associated with the true ν, i.e., ν^*. A graphical illustration is shown in Figure 3.4, where circles (outliers) and triangles (majority) represent two classes that are linearly nonseparable in a 2-D feature space. There are also three hyperplanes represented in the kernel space. These hyperplanes correspond to $\nu = \{\nu_{min}, \nu^*, \nu_{max}\}$. The actual values of ν vary between ν_{min} and ν_{max}. The three oval regions correspond to feature vectors lying in different regions with respect to the optimal hyperplane. Region III corresponds to those samples lying to the right of the hyperplane with $\nu = \nu_{max}$ and so mostly contains the majority data. Region II corresponds to the region around the optimal ν value i.e. ν^* and so has a mixture of outlier and majority data. Region I is composed mostly of outliers and corresponds to the region lying to the left of the hyperplane having $\nu = \nu_{min}$.

SVM score is defined as the distance of a particular data sample in kernel space from the hyperplane. Feature vectors lying near the origin have negative SVM score and those lying to the right of the hyperplane have positive scores. SVM score for a particular sample x in kernel space is calculated by,

$$SVMScore(x) = (x.w) - \rho. \tag{3.2}$$

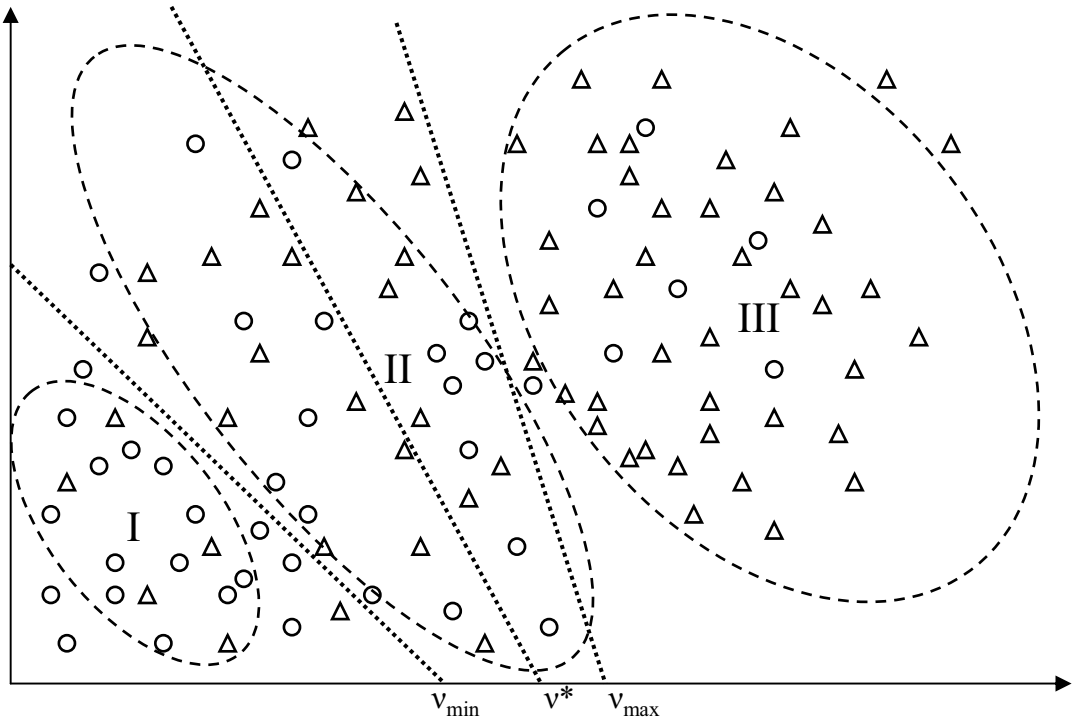


Figure 3.4: Majority (triangles) and Outliers (circles) are represented in kernel space. The hyperplanes formed by varying values of $\nu \in \nu_{min}, \nu^*, \nu_{max}$ are shown. The three oval regions named I,II,III show the sampling areas.

The method using Equation 3.1 may not be accurate because there are always some misclassified samples involved in the computation due to the linear non-separability. On the other hand, region I includes outlier samples with large negative SVM scores, and region III contains majority samples with large positive SVM scores. The samples in regions I and III can be almost for sure correctly classified when $\nu \in [\nu_{min}; \nu_{max}]$. Thus if we use samples in regions I and III as outlier and majority training samples for SVM, the robust classification results that are insensitive to the variations of ν values could be obtained.

3.5.3 Proposed ν-insensitive Approach

Given a test CRP tract X of N samples, we assume that the majority of X is compliant, i.e., $\nu^* < 0.5$. After OCSVM classification, we sort all data samples in

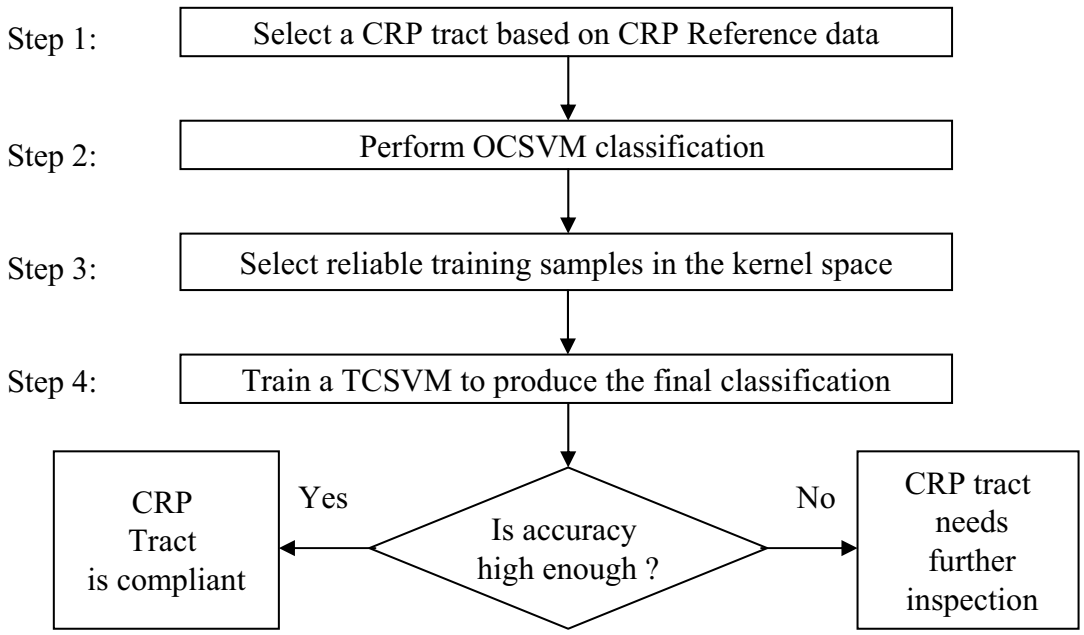


Figure 3.5: Flowchart of Method II.

the majority and outlier classes according to the their SVM score magnitudes from the largest to the smallest. $X_M = \{x_m^i, i = 1, ..., N_+\}$ and $X_O = \{x_o^i, i = 1, ..., N_-\}$; denote the sorted majority and outlier data sets, respectively, where $l = N_+ + N_-$. We define X_M^t and X_O^t as the majority and outlier training sets for SVM, which can be constructed as follows,

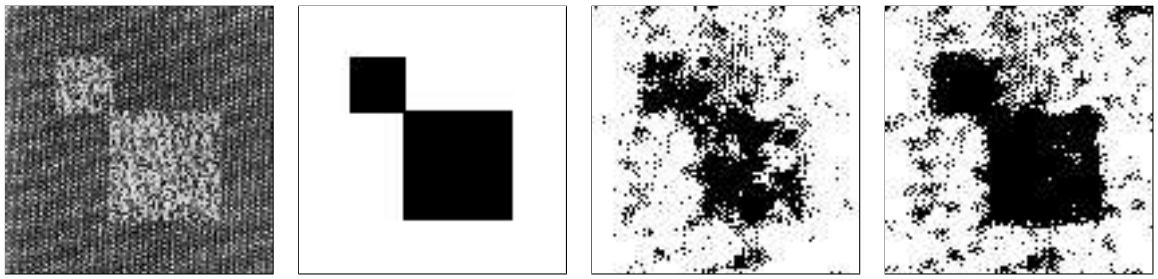
$$\begin{aligned} X_M^t &= \{x_i^m | i = 1, ..., 0.45l\} \\ X_O^t &= \{x_j^o | j = 1, ..., (1-\nu)N_-\}. \end{aligned} \tag{3.3}$$

Since $\nu^* < 0.5$, we use $0.45l$ samples in X_M with largest positive SVM scores as majority training samples (e.g., region III in Figure 3.4). The number of outlier training samples (e.g., region I in Figure 3.4) is set to be $(1-\nu)N_-$. If we choose small ν, small N_- will result. Then most samples in X_O could be true outliers, and we can use most of them for SVM training. On the contrary, if we choose large ν, large N_- will result. X_O may mistakenly contain some majority samples, and we use a small portion of samples in X_O with the largest negative SVM scores.

In practice, X_M^t and X_O^t might not be perfect, and there are still some misclassified training samples for both the majority and outlier classes. In order to

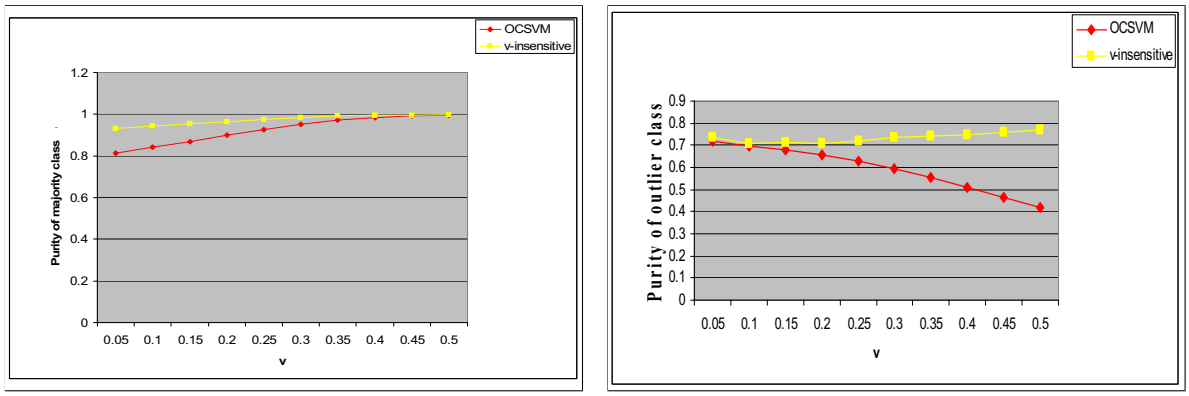
further reduce the side-effect of misclassified training samples, a large margin size is preferred in the SVM, which requires small C value in Equation 2.18.

3.5.4 Experimental Demonstration



(a) (b) (c) (d)

Figure 3.6: Experimental demonstration of the proposed ν-insensitive method based on a synthetic mosaic. (a) mosaic. (b) Ground truth(25% outliers). (c) OCSVM result with $\nu = 0.25$, 85.18% accuracy. (d) The result of the proposed method with $\nu = 0.5$, 84.32% accuracy.



(a) (b)

Figure 3.7: Simulation results on synthetic mosaic. Values are in the range of 0 to 1, where 1 indicated 100% purity. (a) Purity of majority training samples vs. ν. (b) Purity of majority training samples vs. ν.

Specifically, the method proposed in [22] is used to represent the texture pixels in a 25-dimension feature space, which is derived from pixel intensities within a $7X7$ window. The OCSVM is first tested with different ν values ranging from 0.05 to 0.5, and RBF kernel is used with $\gamma = 0.000001$. D_ν achieves the largest value when estimated $\hat{\nu} = 0.25$, which is consistent to true ν^*. Based on the OCSVM results, we

calculate the purity (precision) of the outlier and majority classes regarding different ν values as shown in Figure 3.7 (a) and (b). It is seen that when ν changes from 0.05 to 0.5, the purities of both majority and outlier classes vary considerably. The proposed ν-insensitive method can select sufficient and reliable training samples with higher purity . This leads to robust and ν-insensitive classification results. The best classification accuracy 85.18% of OCSVM with $\nu = 0.25$ is illustrated in Figure 3.6. When testing the proposed method, RBF kernel is also used for SVM with $\gamma = 0.000001$. Even when $\nu = 0.5$, which deviates from true ν^* significantly, we still obtain the similar classification performance (i.e., 84.32%) as the OCSVM that requires many attempts, as shown in Figure (3.6(c) and (d)). This simulation shows the effectiveness and efficiency of the suggest ν-insensitive method.

3.5.5 Experimental Setup

Principal component analysis (PCA) is applied to reduce feature redundancy and to preserve 97% variation of the original images and their texture information. After PCA, around 27 data layers are used in this simulation. A specific software LIBSVM [2] is used to implement OCSVM and SVM in this work. In the OCSVM, the Gaussian kernel with $\gamma = 0.000001$ is chosen according to the cross validation results. The ν value is varied from 0.05 to 0.5 with an interval of 0.05. In the SVM, we select $C = 0.5$ and a Gaussian kernel with $\gamma = 0.01$.

3.5.6 Simulations and Discussions

Simulations are performed on six CRP tracts extracted from Texas County (Section 1.4). In each CRP tract, we also deliberately add some non-CRP regions near to CRP boundaries in order to test the robustness of the proposed methods. In this work, Method-I needs 10 times of OCSVM training and 1 time of SVM training, while Method-II trains both OCSVM and SVM only once, saving more than 80% computational load. The simulation results are shown in Figure 3.8 and Table 3.2. As we can see, the classification performances of both OCSVM and Method-I vary

significantly as ν changes, while Method- II is not sensitive to the variation of ν value with much less standard deviation of the classification accuracy.

Table 3.2: Standard Deviations of the classification accuracy.

CRP Tract	1	2	3	4	5	6
OCSVM	5.19	11.81	2.79	4.45	15.08	11.03
Method-I	7.65	20.97	6.16	5.56	17.34	18.47
Method-II	2.07	3.25	1.19	0.94	5.01	4.06

We illustrate the CRP classification results in Figure 3.9, where five rows from top to bottom refer to, respectively, Landsat images, the CRP reference data (CRP in gray and non- CRP in black), the OCSVM classification results, the results of Method-I where
hatnu is estimated from Equation 3.1, and the results of Method-II where $\nu = 0.4$ (there is no significant change with different ν values). Moreover, the percentage of non-CRP areas according to the CRP reference data (P_{nc}), the percentage of non-CRP areas detected by Method-II (P^*_{nc}), as well as the their differences ($P^*_{nc} - P_{nc}$) are computed for each CRP tract and listed in Table 3.3.

Table 3.3: Non-CRP Percentages(%) Comparison.

CRP Tract	1	2	3	4	5	6
P_{nc}	33.6	29.8	33.7	21.3	9.3	3.7
P^*_{nc}	27.7	25.6	27.6	29.7	32.8	29.8
$P^*_{nc} - P_{nc}$	-6.1	-5.8	-7.9	+8.4	23.5	26.1

In tracts 1, 2, 3, and 4, P^*_{nc} is relatively consistent with or even lower than P_{nc}. Manual inspection further manifests that the CRP areas in tracts 1, 2, 3, 4 have good compliance with respect to the CRP reference data. However, the non-CRP areas in tracts 5 and 6 are have been over detected. This implies that there could be compliance issues in tracts 5 and 6. As observed from the Landsat images in the first row of Figure 3.9, there exist some active cultivation areas (darker areas) in those two tracts, which were previously registered as CRP in the reference data. Therefore tracts 5 and 6 need further detailed inspection. Moreover, there are also

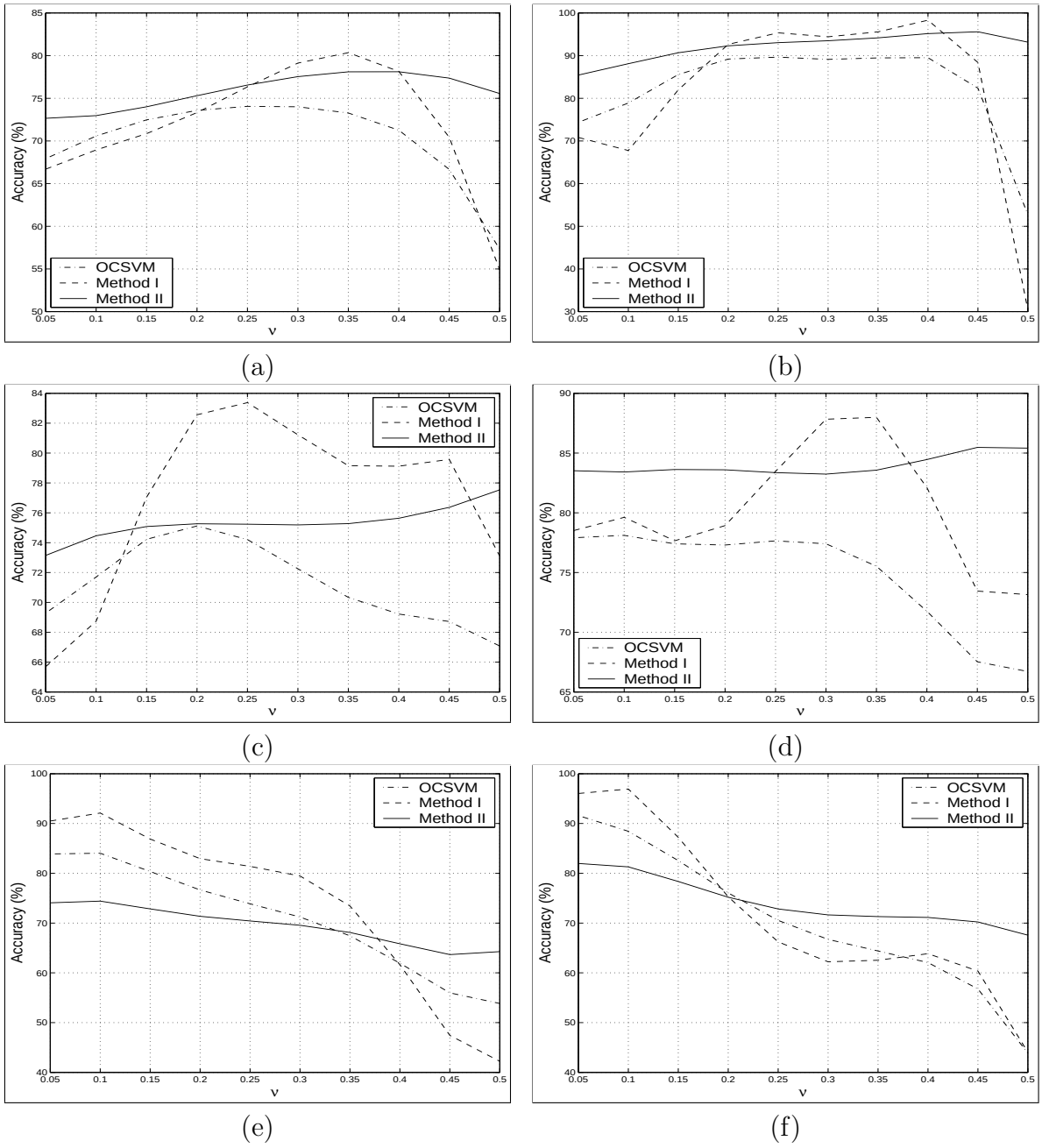


Figure 3.8: The plots of classification accuracy vs. ν for the three methods in six tracts: (a) tract1, (b) tract 2, (c) tract 3, (d) tract 4, (e) tract 5, (f) tract 6.

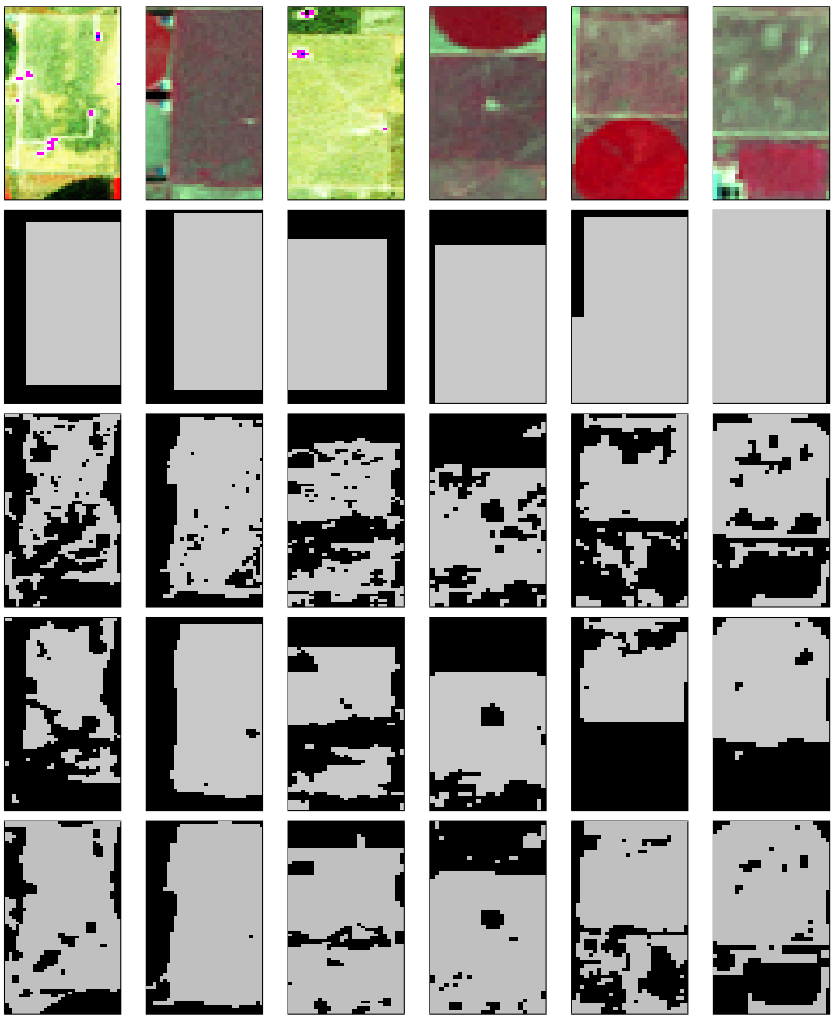


Figure 3.9: (a)-(f) are the simulations of the six test tracts. The rows refer to the original Landsat images(June 2000), CRP Reference data, OCSVM results, Method-I results and Method-II results respectively.

some man-made buildings in tracts 1, 3, 4, which can be clearly detected by Method I and Method-II as well. Nevertheless, only non-CRP percentage values may not provide sufficient information for compliance monitoring, and additional analysis of the CRP classification maps (the last row of Figure 3.9) may be necessary.

In this work, we also found some limitations of our previously proposed Method-I. Largest D_ν is not necessarily related to the best ν or true ν^*. This fact indicates that CRP and non-CRP are not clearly separated even in the high dimensional feature space mapped via RBF kernel. For example, in tract 2, D_ν has the largest value when $\hat{\nu} = 0.4$, while the best OCSVM result with highest classification accuracy is obtained when $\nu = 0.25$ which is close to true ν^*.

3.6 Summary

Two methods have been proposed for CRP compliance monitoring. Method-I relies on distance estimation between the majority and the outliers in the kernel feature space followed by some spatial processing to get reliable training samples. Method-II relies on the natural clustering of data in the kernel feature space to decide on better training samples. The percentage of CRP areas identified imply if the CRP tract is compliant or not.

Simulation results show that both methods can provide useful guidance for effective CRP compliance monitoring. It has been found that Method-II has a more robust classification performance. Also Method-II is more computationally efficient than Method-I. Both the proposed algorithms could be applied to other compliance monitoring problems.

Chapter 4

CRP MAPPING

4.1 Introduction

Existing CRP reference data provided by NRCS is old and has errors due to misalignment of the CRP tracts. Currently CRP maps are developed based on information provided by farmers upon enrollment into the program and by manual delineation of aerial photographs. These maps are needed for reference purposes and for various assessment activities. So it is necessary to develop methods to periodically update CRP maps based on reliable training samples to rectify some locality errors and spatial misalignment of CRP tracts in the reference data. CRP mapping is very different compared to traditional Land Use Land Cover (LULC) applications. CRP mapping is a complex classification problem where both CRP and non-CRP areas are composed of various cover types having highly overlapped clusters in the spectral space of the satellite imagery. Also CRP mapping is an uneven classification task where the CRP tracts amount to less than 10% of the total study area.

4.2 Proposed Approach

Our method is based on Landsat data and CRP ground data (Section 1.4). This process however, is not easy because there is a huge amount of non-CRP regions which have to be avoided during classification. In fact the region under study has only less than 10% of CRP areas (even though it is supposed to have the highest concentrations among Oklahoma's counties). So we require a one-class classifier that

is trained on a particular CRP cover type to find out other instances of the same cover type. However a major problem is the case where a non-compliant CRP tract is use for training. In this case other non-CRP areas may also be classified as CRP. So it is necessary before training to make sure that our classifier is trained only on reliable CRP tracts. Many of the grass species are related. A simplification that can be done is to combine data from multiple but related grass types while training. This can lead to a higher accuracy in classification and also reduce the computational time due to reduced number of Support Vectors. Finally all the classified regions have to be combined to produce the final CRP map.

4.2.1 Pre-clustering of CRP Cover Types

This problem can be stated as determining the optimal number of clusters for an unsupervised classification problem. This has been an ongoing research for several years. Various cluster indices and validity measures have been proposed in literature [18], [19], [14], etc; regarding the selection of an optimal number of clusters. A new approach to this problem has been discussed in [12]. This approach is based on the representations of patterns in kernel spaces projected by Gaussian kernels. Now the elements in the projected space can be represented by a $N \times N$ symmetric kernel matrix K.

$$K = K_{ij} \ ; \ i = 1, ..., l \ ; \ j = 1, ..., l,$$

$$\text{where: } K_{ij} = k(x_i, x_j) \equiv \Phi(x_i).\Phi(x_j) \text{ and } K_{ij} = K_{ji}.$$

This matrix K consists of the dot product distance between different patterns in the kernel space. So it is similar to an adjacency or proximity matrix. Thus this matrix will have a block diagonal structure when there are definite groupings or clusters within the patterns. The eigenvectors of a permuted matrix are the permutations of the original matrix and therefore an indication of the number of clusters can be obtained from the eigenvalue decomposition of the kernel matrix.

Eigenvalue decomposition of the kernel matrix gives $K = U\Lambda U^T$. Where the

diagonal matrix Λ contains the eigenvalues denoted as λ_i. The columns of U matrix contain the individual eigenvectors u_i. So we can write,

$$\begin{aligned} 1_l^T K 1_l &= 1_l^T \{\sum_{i=1}^{l} \lambda_i u_i u_i^T\} \\ &= \sum_{i=1}^{l} \lambda_i \{1_l^T u_i\}^2. \end{aligned} \tag{4.1}$$

In Equation 4.1, 1_l is a $l \times 1$ vector with elements of value $1/l$. The final form in Equation 4.1 indicates that if there are K distinct clustered regions within the l patterns then there will be K dominant terms $\lambda_i\{1_l^T u_i\}^2$ in the summation.

4.2.2 Combining multiple OCSVM's

Combining different one-class classifiers has been discussed in [36] and [34]. But these describe combining different varieties of one-class classifier trained on the same patterns. Our problem however deviates from the general case. We want to train each of the different One-Class Classifiers on a different class of patterns. So hence it is more similar to the case described in [23], where an Image Database Retrieval problem is discussed. Figure 4.1 depicts that, it is not justified to combine unrelated cover types as is shown using a single hyperplane H3 created by a single OCSVM. So multiple OCSVM's are needed to increase CRP sensitivity.

In our method each class of grass species are trained individually so that their unique representation (support vectors) is obtained. Now this representation is used to classify the entire study area. Also some grass species overlap in kernel space so it is justifiable to combine them. Combining diverse grass species is not advantageous because the similar number of Support Vectors will be maintained offering no reduced complexity for the classifier and no improvements in classification time and performance; however a disadvantage is that larger amount of non-CRP data will be classified as CRP (Figure 4.1).

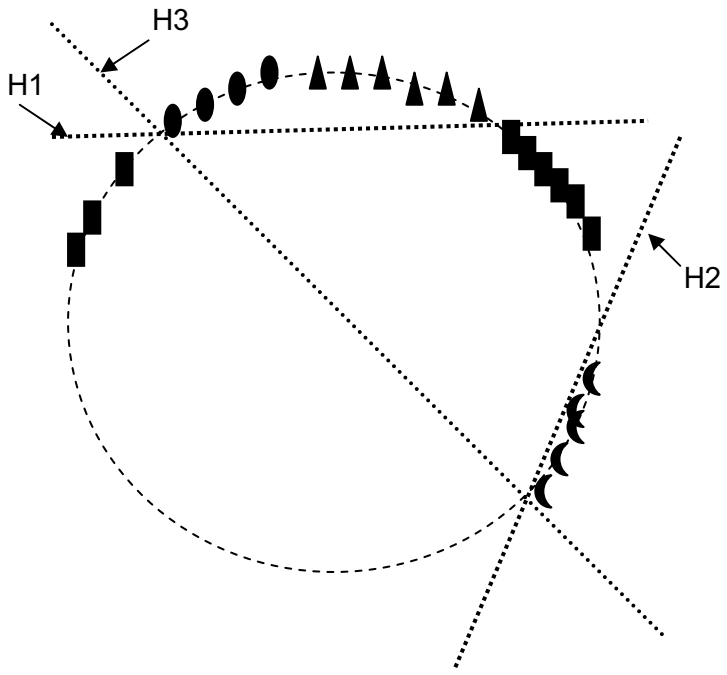


Figure 4.1: Kernel space representation of multiple OCSVMs using the Gaussian kernel (refer Figure 2.9). Here oval, triangle and crescent shaped objects represent feature vectors of different CRP cover types, rectangles depict feature vectors of non-CRP covers. The oval and triangular CRP covers are related. H1, H2 and H3 are different hyperplanes subtended by different OCSVMs. It can be seen that using H1 and H2 separately avoids misclassification of non-CRP data as CRP, however using H3 alone some non-CRP data is classified as CRP.

4.3 Simulations and Discussions

The data used for simulations is a clip from the multi-temporal Landsat data of Texas county of size 552×523 pixels where each pixel has an area of $30m \times 30m$. Simulations have been done using different OCSVM's combined together to form a final result. OSCVM implementation in [2] was used for the simulation. As this mapping is specifically for CRP regions we are interested in having higher accuracy for CRP areas even if it means that more non-CRP areas will be incorrectly classified. So for the OCSVM, the Gaussian kernel was used and the γ value was set to 1 to achieve tighter classification boundaries. The ν parameter was set at 0.1 to avoid some highly likely outlier data within the CRP training data. CRP tracts which

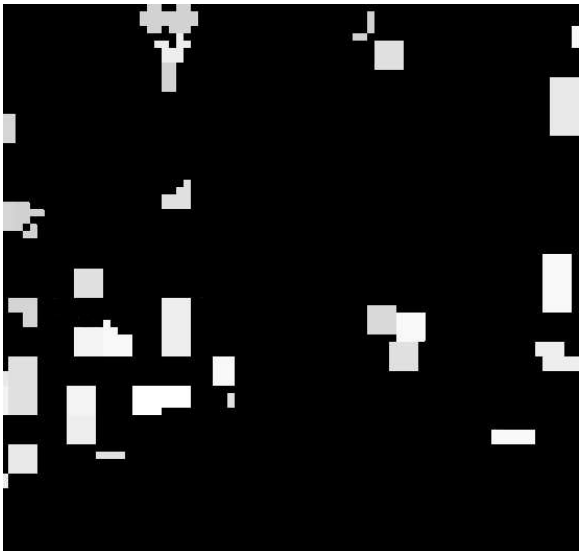


Figure 4.2: Ground data for the simulations. Black regions are non-CRP and grey regions are CRP.

are totally non-compliant were excluded. A single OCSVM classifier's training data is formed by combining together grass species data which had only one dominant term, as in the Equation 4.1. This process was done by ordered combination of the different CRP cover types till the minimum number of classifiers were obtained.

The classification and training were repeated 20 times using different training datasets (developed using different initializations for the random sampling procedure) and the average values are shown in the three Tables (4.1, 4.2 and 4.3) which are used to display the results. The cover types given in normal lettering are the different grass species which together compose the CRP cover type. The non-CRP cover type represents the accuracies for all other cover types not included in the CRP class. A weighted accuracy for all CRP cover types is given as the CRP cover type. The total accuracy for all the data (CRP + non-CRP) is given as the overall cover type. The percentage values on the first row (column headings) of the tables represent the amount of total CRP data used for training the OSCVM classifiers. This was done using a random sampler to obtain the required sampling size. Count (a column heading) specifies the number of pixels in the testing data belonging to each cover type. All numerical values in the table expect for the column count are in percentage.

Table 4.1: Accuracy rates without using grass combinations.

CoverType	**Count**	**40% sampling**	**70% sampling**
non − CRP	265686	93.77%	89.16%
Old World Bluestem	585	74.27%	87.34%
Plains Bluestem	833	36.18%	68.90%
WW Spar	1029	49.67%	75.99%
Plains Bluestem (1986)	924	62.18%	83.63%
Granada (1986)	336	40.66%	81.41%
Caucasian (1987)	784	66.83%	84.22%
Plains Bluestem (1987)	4765	71.93%	87.42%
Plains (1988)	2422	55.16%	81.76%
Plains (1989)	3843	56.49%	78.31%
WW Spar (1989)	1708	58.94%	82.12%
Old World Bluestem (1990)	4409	68.82%	85.63%
Native Mixture (1990)	1372	59.33%	81.37%
CRP	**23071**	**61.61**%	**82.01**%
Overall	**288696**	**90.76**%	**91.19**%

Table 4.2: Accuracy rates using OCSVMs trained on grass combinations.

CoverTypes	**Count**	**40% sampling**	**70% sampling**
non − CRP	265686	90.89%	86.12%
Old World Bluestem	585	77.76%	88.21%
Plains Bluestem	833	36.65%	69.87%
WW Spar	1029	52.93%	79.97%
Plains Bluestem (1986)	924	64.97%	82.98%
Granada (1986)	336	42.18%	79.16%
Caucasian (1987)	784	67.48%	84.56%
Plains Bluestem (1987)	4765	73.71%	88.56%
Plains (1988)	2422	56.21%	82.08%
Plains (1989)	3843	59.67%	80.45%
WW Spar (1989)	1708	64.82%	83.51%
Old World Bluestem (1990)	4409	71.38%	86.31%
Native Mixture (1990)	1372	61.13%	81.06%
CRP	**23071**	**64.32**%	**83.32**%
Overall	**288696**	**88.76**%	**85.89**%

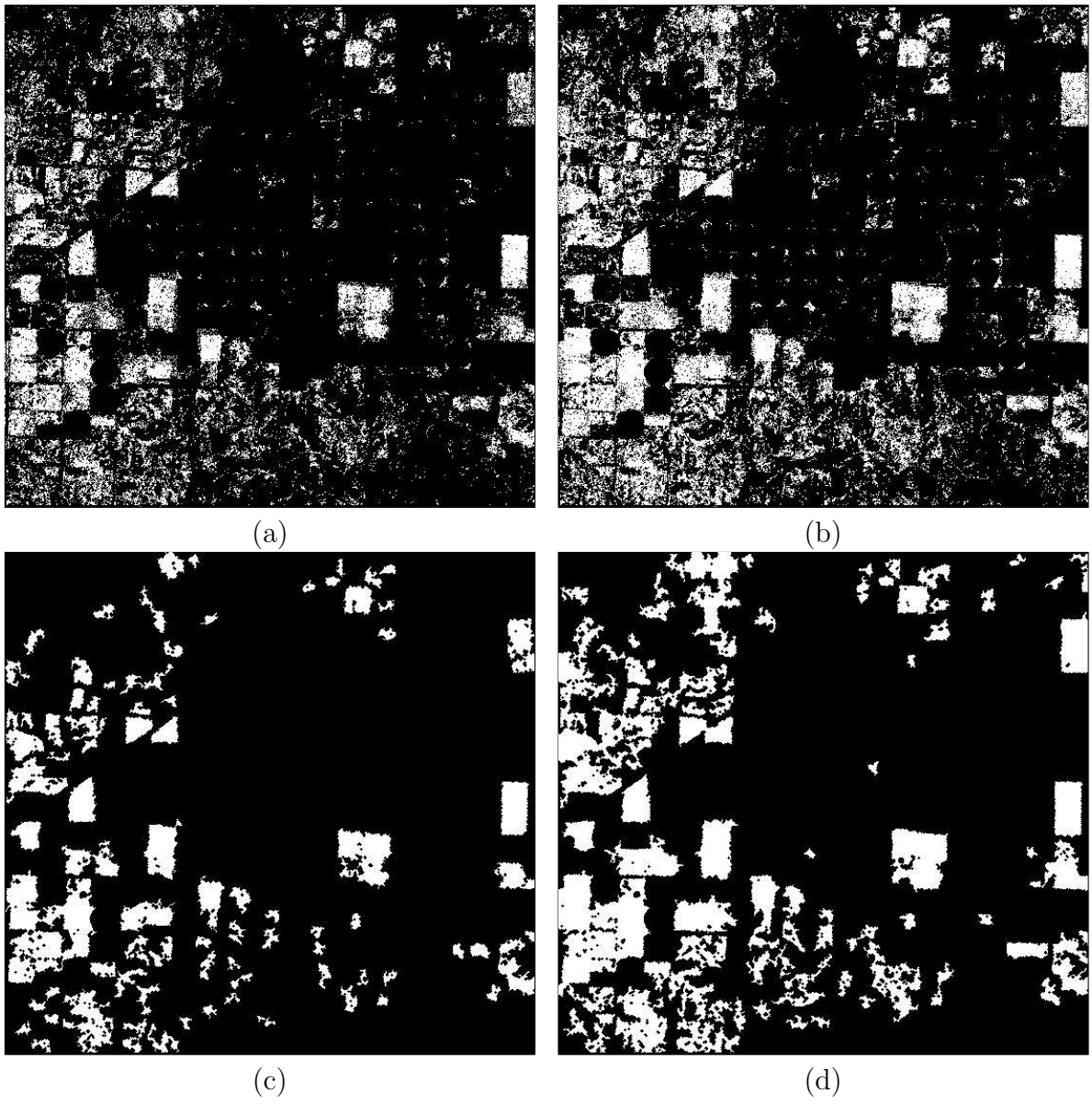


Figure 4.3: The classification results after combining multiple OCSVM's for different training data sets, (a) 40% sampling size with pre-clustering and without post-processing, (b) 70% sampling size with pre-clustering and without post-processing, (c) 40% sampling size with pre-clustering and after post-processing, (d) 70% sampling size with pre-clustering and after post-processing.

Table 4.3: Accuracy of OCSVM Classifier trained with pre-clustered data after morphological processing.

CoverTypes	**Count**	**40% sampling**	**70% sampling**
non − CRP	265686	93.56%	87.41%
Old World Bluestem	585	89.97%	97.28%
Plains Bluestem	833	20.55%	90.61%
WW Spar	1029	64.43%	96.59%
Plains Bluestem (1986)	924	68.25%	95.02%
Granada (1986)	336	46.72%	98.80%
Caucasian (1987)	784	85.02%	94.00%
Plains Bluestem (1987)	4765	82.33%	93.98%
Plains (1988)	2422	70.93%	97.92%
Plains (1989)	3843	66.18%	96.57%
WW Spar (1989)	1708	70.67%	96.89%
Old World Bluestem (1990)	4409	74.56%	95.97%
Native Mixture (1990)	1372	71.92%	94.48%
CRP	**23071**	**71.22%**	**95.56%**
Overall	**288696**	**91.44%**	**88.06%**

Table 4.1 shows the classification accuracy without using the pre-clustering process. So here a OCSVM is trained for each cover CRP cover type individually and the testing results for the full dataset are combined together. Table 4.2 shows the classification accuracy using the pre-clustering process. The classification maps obtained after combining the different OSSVMs in this case is given in Figure 4.3. Here seven OSCVM classifiers are trained. The different CRP Cover types that are grouped together are,

- Old World Bluestem.
- Plains Bluestem.
- Caucasian (1987).
- Plains (1988).
- Granada (1986) and Native Mixture (1990).
- WW Spar, Plains (1989) and Old World Bluestem (1990).

- Plains Bluestem (1986), Plains Bluestem (1987) and WW Spar (1989).

As can be seen the pre-clustering improves the accuracy of the classification by an average of 2-3% but the accuracy improvement for individual grass species is even higher. Table 4.3 gives the accuracies after doing morphological processing on the outputs produced by the OCSVMs trained with pre-clustered data. The classification maps obtained after combining the different OSSVMs in this case is given in Figure 4.3. Morphological processing is used to improve the visual appearance of the map as well as increasing the classification accuracy. Initially all elements in the map having the number of connected components less than 50 were removed to get rid of noise and misclassified regions. Then morphological closing operation was used to close small holes within the connected regions. As can be seen from the three Tables (4.1, 4.2 and 4.3) the overall accuracy reduces as we increase the sampling size. This is due to the reduction in non-CRP detection accuracy. However what we are interested is in getting higher amount of CRP detection accuracy which is obtained when increasing the sampling size. Another factor to be kept in mind is that the CRP reference data we currently have may not be completely accurate so these accuracies serve only as guidelines. To get the perfect accuracy for our method it is necessary to obtain true data by conducting field trips when the satellite images were acquired.

The computational time required for the total classification process trained on pre-clustered data is given in Table 4.4. The simulations were done on an Intel P4 2.4GHz system with 1 GB of RAM. As expected there is reduced processing time when the amount of sampling is reduced. When classifying with OCSVMs having pre-clustered data the computing time is reduced as there are less number of OCSVMs to train and classify. Also the number of *Support Vectors* are reduced so more compact representation of the training data is used for classifying the testing dataset. This reduces the classification time.

Table 4.4: Simulation times for one run of the classification.

Sampling Rate	With pre-clustering	Without pre-clustering	Speed Improvement
40%	23 min	30 min	23%
70%	35 min	49 min	28%

4.4 Summary

Conservation Reserve Program mapping has been implemented using multiple-OCSVM's. Kernel space proximity was used as a criterion to combine data from different grass species so as to marginally improve classification performance and decrease computational time. High accuracy rates for the classification have been obtained.

Chapter 5

CONCLUSIONS AND FUTURE WORK

In this thesis, we have studied two specific remote sensing issues related to USDA's CRP program. Specifically, we have proposed two CRP compliance monitoring methods and one CRP mapping technique. Currently, CRP mapping and compliance monitoring are accomplished manually, which is very costly and time-consuming. SVM-based methods using Landsat TM imagery will allow prompt and accurate CRP classification results that are valuable for CRP management and evaluation.

- For CRP compliance monitoring, we have implemented two methods. The first one is a distance estimation based approach that needs to estimate the ν parameter iteratively for one-class SVM (OCSVM). The second is the ν-insensitive method which is does not need the ν estimation. Performance of the ν-insensitive method was more efficient than the first method with better robustness and less computational time. The simulation results are satisfactory considering the complexity of multiple CRP species. However the two methods have to be validated based on field data collection. At present, it can serve as a guideline for more detailed study into specific CRP plots. As the first work of its kind, it also will serve as a benchmark for future work into this area.

- For CRP mapping we have implemented a method by combining multiple OCSVMs which are trained on different CRP cover types. Also a pre-clustering procedure has been developed to merge different CRP cover types into one class

before OCSVM training. This process has been found to marginally improve classification accuracy and to reduce the classification complexity dramatically. Simulations were based on varying sampling sizes of the training data, and it is shown that high accuracy can be achieved even though the training data was less than 10% of the entire testing data.

Future work on compliance monitoring will have to relax the constraint that the majority of a CRP tract should be compliant, since there may be the case which the majority of a CRP tract is not compliant. So the objective will be to split the data into multiple clusters. This will be based on kernel space distribution of the feature vectors and the OCSVM can be used to detect the clusters by iteratively removing the largest clusters based on inter-cluster and intra-cluster distance. Residual insignificant noisy data will have to be neglected. Knowledge of CRP and non-CRP cover types will help in deciding the cluster memberships. Thus producing a more reliable compliance monitoring result.

Advances in CRP mapping will be to consider group based learning methods to combine the results from multiple OCSVM classifications by varying ν values thus obtaining more accurate classification by using purer raining data. Usage of GIS (Geographic Information Systems) data like slope, elevation, etc; may also improve the classification accuracy as the GIS information for CRP lands should confirm to that of agricultural lands. Thus this will reduce the over-detection of CRP tracts in the areas where the cultivation was never practiced.

Also both compliance monitoring and mapping can be combined into a joint framework where compliance monitoring will be used to obtain reliable CRP training data for the mapping procedure. Field studies need to be conducted to validate the proposed methods. The procedures that we have developed here could be applied to similar problems where knowledge of classification is only available about the classes which are of interest.

BIBLIOGRAPHY

[1] *Conservation Reserve Program.* http://www.fsa.usda.gov/dafp/cepd/crp.htm.

[2] *Libsvm: a library for support vector machines.* 2003. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[3] A. S. Belward. A comparison of supervised maximum likelihood and decision tree classification for crop cover estimation from multitemporal landsat mss data. *Int. J. Remote Sensing*, 2(2):229–235, Dec. 1987.

[4] M. Brown, H. G. Lewis, and S. R. Gunn. Linear spectral mixture models and support vector machines for remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 38(5):2346–2360, September 2000.

[5] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, Jun. 1998.

[6] G. Cherian, X. Song, G. Fan, and M. Rao. Application of support vector machines for automatic compliance monitoring of the conservation reserve program (crp) tracts. *Proc. IEEE International Geoscience and Remote Sensing Symposium*, September 2004.

[7] R. Defries, M. Hansen, J. Townshend, and R. Sohlberg. Global land cover classification at 8km spatial resolution: the use of data derived from landsat imagery in decision tree classifiers. *Int. J. Remote Sensing*, 19(16):3141–3168, 1998.

[8] S. L. Egbert, R. Y. Lee, K. P. Price, M. D. Nellis, and R. Boyce. Mapping conservation reserve program (crp) lands using multi-seasonal thematic mapper imagery. *GeoCarto International*, 13(4):17–24, 1998.

[9] S. L. Egbert, S. Park, K. P. Price, Re-Yang Lee, J. Wu, and M. D. Nellis. Using conservation reserve program maps derived from satellite imagery to characterize landscape structure. *Computers and Electronics in Agriculture*, 37(1-3):141–156, Dec. 2002.

[10] G. M. Foody and A. Mathur. A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(6):1335–1343, June 2004.

[11] M. Friedl and C. Brodley. Decision tree classification of land cover from remotely sensed data. *Remote Sensing Environment*, 61(3):399–409, 1997.

[12] M. Girolami. Mercer kernel-based clustering in feature space. *IEEE Transactions on Neural Networks*, 13(3):780–784, May 2002.

[13] J. A. Gualtieri and R. F. Cromp. Support vector machines for hyperspectral remote sensing classification. *in Proc. SPIE AIPR*, 3584:221–232, 1998.

[14] M. Halkidi, Y. Batistakis, and M. Vazirgiannis. On cluster validation techniques. *Journal or Intelligent Information Systems*, 17(2/3):107–145, 2001.

[15] J. Heald. Usda establishes a common land unit. *ArcUser Online*, March-April 2002. http://www.esri.com/news/arcuser/0402/usda.html.

[16] L. Hermes, D. Frieauff, J. Puzicha, and J. M. Buhmann. Support vector machines for land usage classification in landsat tm imagery. *Proc. IEEE International Geoscience and Remote Sensing Symposium*, pages 348–350, 1999.

[17] C. Huang, L. S. Davis, and J. R. G. Townshend. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4):725 – 749, Feb. 2002.

[18] A.K. Jain and R.C. Dubes. *Algorithms for Data Clustering.* Englewood Cliffs,NJ: Prentice Hall, 1988.

[19] A.K. Jain, M.N. Murty, and P.J. Flynn. Data clustering: A review. *ACM Computing Surveys*, 13(3):264–323, Sept. 1999.

[20] S. S. Keerthi and C-J. Lin. Asymptotic behaviors of support vector machines with gaussian kernel. *Neural Computation*, 15(7), July 2003.

[21] J. Keuchel, S Naumann, M Heiler, and A. Siegmund. Automatic land cover analysis for tenerife by supervised classification using remotely sensed data. *Remote Sensing of Environment*, 86(4):530–541, August 2003.

[22] K. I. Kim, K. Jung, S. H. Park, and H. J. Kim. Support vector machines for texture classification. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(11):1542–1550, Nov. 2002.

[23] C. Lai, D.M.J. Tax, R.P.W. Duin, E. Pekalska, and P. Paclik. On combining one-class classifiers for image database retrieval. *in* Proc. Multiple Classifier Systems, MCS 2002*,J. Kittler, F. Roli (eds.)*, pages 212–221, 2002.

[24] F. Melgani and L. Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8):1778–1790, Aug. 2004.

[25] K-R. Müller, S. Mika, G Rätsch, K. Tsuda, and B. Schölkopf. An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*, 12(2):1443–1471, 2001.

[26] G. Rätsch, S. Mika, B. Scholkopf, and K.-R. Müller. Constructing boosting algorithms from svms: an application to one-class classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1184–1199, Sept. 2002.

[27] F. Roli and G. Fumera. Support vector machines for remote-sensing image classification. *in Proc. SPIE*, 4170:160166, 2001.

[28] B. Schölkopf. *Support Vector Learning*. Ph.D. dissertation, Technischen Universität Berlin, 1997.

[29] B. Schölkopf, S. Mika, C. J. C. Burges, P Knirsch, K-R. Müller, G Rätsch, and A. J. Smola. Input space verses feature space in kernel-based methods. *IEEE Transactions on Neural Networks*, 10(5):1000–1017, September 1999.

[30] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 2001.

[31] M. Simard, S. S. Saatchi, and G. D. Grandi. The use of decision tree and multiscale texture for classification of jers-1 sar data over tropical forest. *IEEE Trans. Geoscience and Remote Sensing*, 38(5):23102321, Sept. 2000.

[32] X. Song, G. Cherian, G. Fan, and M.Rao. A ν-insensitive svm approach for automatic compliance monitoring of the conservation reserve program (crp) tracts. *resubmitted after corrections to IEEE Geoscience and Remote Sensing Letters*, August 2004.

[33] X. Song, G. Fan, and M. Rao. Machine learning approaches for multisource geospatial data classification with application to crp mapping in texas county, oklahoma. *Proc. IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, October 2003.

[34] D. M. J. Tax. *One-class Classification*. Ph.D. dissertation, Technische Universiteit Delft, The Netherlands, 2001.

[35] D. M. J. Tax and R. P. W. Duin. Support vector data description. *Machine Learning*, 54(1):45–66, Jan. 2004.

[36] D.M.J. Tax and R.P.W. Duin. Combining one-class classifiers. *in* Proc. Multiple Classifier Systems, MCS 2001 *,J. Kittler, F. Roli (eds.)*, pages 299–308, 2001.

[37] R. Unnthorsson, T. P. Runarsson, and M. T. Jonsson. Model selection in one-class ν-svms using rbf kernels. 2003. Paper available at http://www.hi.is/ runson/svm/paper.pdf.

[38] V. N. Vapnik. *The Nature of Statistical Learning Theory*. New York: Spinger-Verlag, 1995.

VITA

Ginto Cherian

Candidate for the degree of

Master of science

Support Vector Machines for Conservation Reserve Program (CRP) Mapping and Compliance Monitoring

Major Field: Electrical Engineering

Biographical:

Born in Vellathooval, Kerala, India, on Feburary 22, 1980, the son of O. V. Cherian and Mary Cherian.

Graduated from St. Berchmans College, Changnacherry, India in May 1997. Received Bachelor of Technology degree in Computer Science and Engineering from Mahatma Gandhi University, India, in June 2001. Completed the requirement for the degree in Master of Science Electrical Engineering in December, 2004 from Oklahoma State University.

Worked as a Software Engineer at Win Win Computer Solutions, India from September 2001 to June 2002. Presently working as a Research Assistant in the department of Electrical and Computer Engineering at Oklahoma State University since January 2003.

A Student Member of IEEE since January 2003.

ABSTRACT

Advisor : Dr. Guoliang Fan

This research focuses on two specific remote sensing problems associated with the United States Department of Agriculture (USDA)'s Conservation Reserve Program (CRP). The CRP program seeks to convert highly erodible lands with active crop production to permanent vegetative cover. Specifically, there are two essential needs pertaining to CRP management and evaluation, i.e., CRP compliance monitoring and CRP mapping. Multi-spectral and multi-temporal Landsat TM images are used to generate the data. Compliance monitoring checks if a CRP tract is following the contract stipulations. CRP mapping produces up-to-date and accurate maps of CRP lands based on satellite images. We invoke approaches based on the Support Vector Machine (SVM) to address these two issues where two classes, CRP and non-CRP, are involved for data classification. The SVM is a recently developed supervised classifier aiming at maximizing the margin between two clusters in a projected feature space. The SVM was also adapted to solve one-class problems, i.e., novelty detection. The one-class SVM (OCSVM) has a parameter ν to control the percentage of outliers or minority data.

We propose two SVM-based methods for CRP compliance monitoring. The first method uses OCSVMs trained with different ν values and selects the one producing the maximum margin between clusters in the projected space. Then a two-class SVM (TCSVM) is trained by using the initial OCSVM classification results. The second method only involves the OCSVM training once. Training samples for the TCSVM are selected based on their relative locations to the classification boundary in the projected feature space i.e. SVM scores. The second method is proven more efficient than the first one with similar or better accuracy. We also develop

an OCSVM based algorithm for CRP mapping. Reduction in classifier complexity is achieved by combining different CRP species based on their data distribution in a projected feature space. Multiple OCSVMs are trained on different CRP cover types, and are applied for CRP classification individually. The complete CRP map is obtained by merging the different classifier outputs.

This research further manifests the usefulness of the SVM in remote sensing applications. The proposed compliance monitoring and mapping tools will be valuable for CRP management and evaluation.