Support Vector Machines in Classification

Vivek Anand (21103276) [1]

[1] Geoinformatics, Civil Engineering Department, Indian Institute of Technology, Kanpur

Abstract:
Classification in satellite images is an important task as it gives us information about all the objects and classes that are present in an area because the knowledge of the same finds application in a wide variety of areas including but not limited to disaster management, and Earth observation, environmental monitoring, etc. Support Vector Machine (SVM) is one such supervised classification algorithm that finds use in classifying a wide variety of remotely sensed images. This term paper aims at providing an overview of how SVM which is initially a binary classification technique can be used to solve multiclass classification problems which are present in the remotely sensed images.

Keywords:
SVM, Supervised Classification, Satellite Images

1. Introduction:
Classification of images is an important field whether it is remote sensing, pattern recognition, or image analysis. In a number of cases, the main pivot of the analysis is the classification itself. The classification of land use using remotely sensed data, for example, yields a map-like image as a result. For, digital images, the image classification is an important method of analysis. Different classification algorithms are available at present based on the purpose of analysis.

New classification techniques, such as decision trees and ANNs have become part of the mainly used image classifiers as the technology has progressed. These new strategies have been compared to traditional ones in studies, and they have been found to have enhanced classification accuracy[1]. Despite this, there is still a lot of room for improvement in terms of accuracy, as well as a strong desire to extract as much land cover information as possible from remotely sensed data [2]. As a result, research into novel classification algorithms has increased, and support vector machines (SVMs) have recently grabbed the remote sensing community's interest [3].

2. Support Vector Machines
Support Vector Machines , SVMs has its core from the Statistical Learning Theory [4]. They've been used in a variety of machine vision applications, including character, handwriting, digit, and text recognition [5], and, more recently in the classification of the satellite images . SVMs are considered to be robust just like Artificial Neural Networks (ANNs) and other nonparametric classifiers, [6] SVMs work by using a kernel function to nonlinearly project the training data in the input space to a higher dimension feature space. This produces a linearly separable dataset that a linear classifier can easily separate. This method allows remote sensing datasets to be classified, even if they are nonlinearly separable in the input space. In the case of high-dimensional feature spaces the classification model results in over-fitting in the input space; however, over-fitting in SVMs is regulated by the structural risk minimization concept [7].

By increasing the distance(or margins) between the data points and the decision border, the empirical risk of misclassification is drastically reduced[8]. In practise, this requirement is relaxed to the minimization of a cost factor that includes the classifier's complexity as well as the degree to which marginal points are being misclassified. A margin of error parameter which is designated by C is modified using cross-validation techniques which manages the balance between these factors [9]. Kernels are functions that are used to project data from input space to feature space. Examples include polynomial, radial basis functions), and quadratic functions. Prior to classification, each function has its own set of parameters that must be determined, which is commonly done through a cross validation procedure.

SVMs are inherently binary classifiers by nature but there are ways for applying them to multiclass problems connected with remote sensing research. The One-Against-One (1A1) and One-Against-All (1AA) strategies are two popular methodologies.
A support vector classifier in a two-class scenario tries to find a hyperplane that minimises the distance between members of each class and the optional hyperplane. A two-class classification problem can hence be explained as: For Example, there are N training sets that can be given by the set pair {(x , y),i = 1,2,3......,N)} I, with i being the class label of value ±1 and i y ∊ m where feature vector with m components. The classifier is given by the function f (y;α) → x with α , the parameter factors of the classifier. The figure 1 shows the Maximum-margin hyperplane and margins for an SVM trained with samples from two classes. Samples on the margin are called the support vectors.

The SVM algorithm finds an optimal separating hyperplane in which:1) samples with labels ±1 are located on each side of the hyperplane; 2) support vectors are the distance between the nearest vectors to the hyperplane in each side of maximum, and the distance is the best margin.. The hyperplane is given by the equation by w.y +b = 0 where (w,b) are the parameter factors of the hyperplane. The vectors that are not on this hyperplane is given by w.y +b > 0 and let the classifier to be given as f (y;α) = sgm(w.y + b). The support vectors lie on two hyperplanes, which are parallel to the optimal hyperplane, of equation w.y +b = ±1.The maximization of the margin with the equations of the two support vector hyperplanes contributes to the following constrained optimization problem min1/2 $||w||^2$ with x_i (w.y +b) ≥1,i =1,2,.......,M

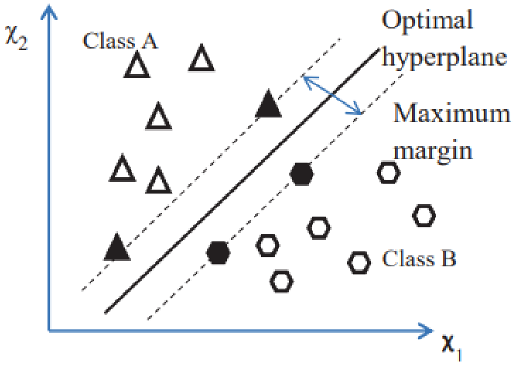

Figure 1: Hyperplane in SVM

3. SVM Multiclass Strategies:

SVM classification is essentially a binary (two-class) classification technique that must be modified to handle multiclass tasks in real-world situations, such as deriving land cover information from satellite images. The 1A1 and 1AA approaches are two typical methods for enabling this adaptation. The 1AA technique (Melgani and Bruzzone, 2004) divides an N-class dataset into N two-class cases and is the first and most widely used SVM multiclass strategy If, for example, the classes which are of our use in an image are water, vegetation, and built-up regions then the classification will be carried out by classifying water against non-water areas (vegetation and built-up areas) or vegetation against non-vegetative areas (vegetation and built-up areas) (water and built up areas). The 1A1 method, on the other hand, entails building a machine for each pair of classes, yielding N(N-1)/2 classifier. When each classification is applied to a test point, the winning class receives one vote, and the point is labelled with the class that received the most votes. This strategy can be tweaked to give the voting process more weight. According to machine learning theory, the 1AA strategy has the disadvantage that its performance can be harmed by unbalanced training datasets[10]; nevertheless, the 1A1 approach is more computationally expensive because the results of more SVM pairs must be computed.

As evident from the above explanation the one-vs-all and one-vs-one methods simply reduces the multiclass classification problem to multiple binary classification problem.

4. Dataset Used

The data used in this study is the Pavia University hyperspectral data which contains 610 × 340 pixels with 115 spectral bands in the range from 0.43 to 0.86 μm with a spatial resolution of 1.3 m/pixel. There are nine ground truth classes

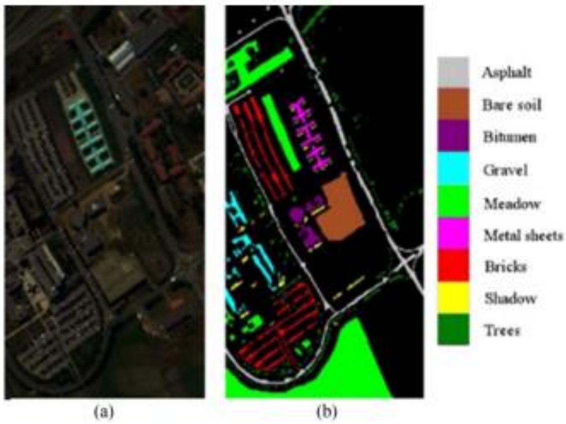

Figure 2: Pavia University HSI: (a) composite color image, (b) ground truth.

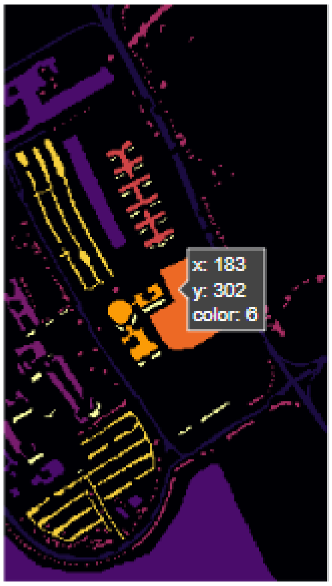

Figure 4: Ground Truth Data

5. Methodology:

The section will outline the procedure followed to classify the hyperspectral data using Support Vector Machines. The following steps were performed to implement SVM in Python programming language:

5.1. Processing the Data:

The HSI data and ground truth data were analysed and inspected, and the bad bands were removed and the remaining 103 bands were used for the processing. All the pixel values from the 103 bands were extracted and stored into a data frame and the ground data which contained the class label was also appended in the same data frame.

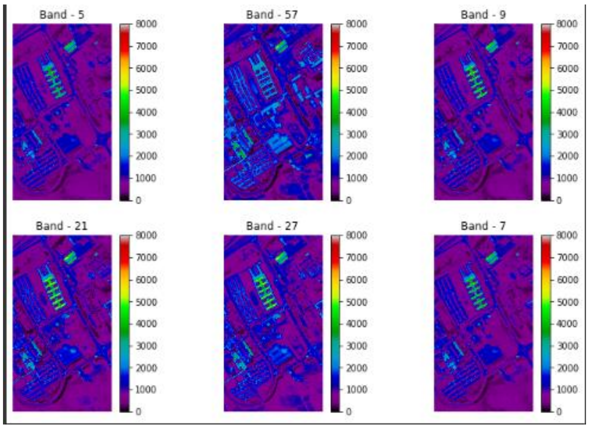

Figure 3: Inspecting Individual Band in the Image.

	band1	band2	band3	band4	band5	band6	band7	band8	band9	band10	...	band95	band96	band97	band98	band99	band100	band101	band102	band103	class
0	647	499	464	371	291	319	365	322	296	305	...	3416	3335	3256	3226	3205	3210	3221	3238	3250	0
1	604	546	527	455	378	377	336	314	324	329	...	2583	2536	2509	2511	2501	2468	2442	2464	2528	0
2	621	746	556	360	285	300	274	276	332	319	...	2288	2242	2193	2282	2225	2266	2308	2345	2361	0
3	637	592	482	556	508	284	178	198	193	224	...	2675	2609	2541	2551	2571	2569	2573	2620	2644	0
4	625	560	480	360	377	341	264	208	231	266	...	2392	2344	2289	2328	2415	2424	2386	2431	2456	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
207395	574	594	504	411	468	443	378	342	352	434	...	2150	2157	2157	2169	2151	2108	2077	2076	2136	0
207396	485	644	709	582	504	494	434	373	391	421	...	2264	2295	2282	2265	2275	2293	2274	2274	2258	0
207397	593	387	428	461	470	392	331	383	439	355	...	2393	2407	2433	2413	2373	2389	2382	2407	2423	0
207398	593	751	655	458	503	560	482	412	404	454	...	2340	2333	2332	2288	2251	2280	2312	2308	2289	0
207399	689	720	449	402	409	410	424	416	395	319	...	2496	2502	2483	2488	2498	2455	2416	2447	2485	0

207400 rows × 104 columns

Figure 5: Data Frame Containing the Pixel values and Class information

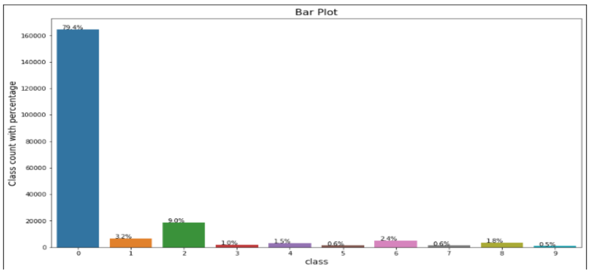

Figure 6: Class wise Pixel Count

5.2. Feature Extraction using PCA:

As the hyperspectral data has very dimensions and it will be computational very expensive to work with it, Principal Component Analysis was applied on the HSI data to find out the principal components which contains the maximum information of the given data. Although on inspection, the first three principal components contained the maximum information but for the sake of inclusion, first five principal components were used to carry out the further processing.

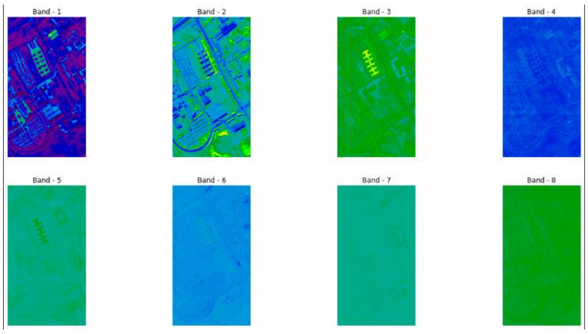

Figure 7: Principal Components

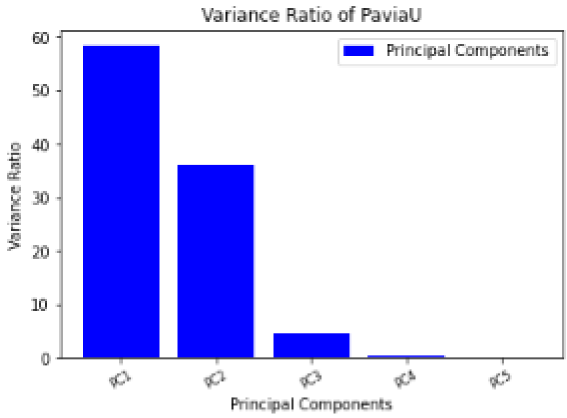

Figure 8: Explained Variance Ratio of the Principal Components

5.3. Classification by SVM:

The principal components were added to a data frame and the corresponding class labels were added and the data was split into training and testing sets. 70% of the data was used to train the model and 30% was used to test the model. The model was then trained and then tested using the test data and the accuracy of the model was calculated.

6. Results and Discussion:

SVM model was trained and used as a classification algorithm to classify the HSI data and various accuracy assessment of the trained model was done and they were visually inspected. Three SVM models were trained and the results were obtained for the same i) SVM with Linear Kernel ii) SVM with Radial Bias Function (RBF) Kernel iii) Multiclass SVM which I have tried to develop and implement from the developed understanding of the algorithm.

6.1.SVM with Linear Kernel:

Using this algorithm, an accuracy of 79.71 % was achieved. The following figure 9 shows the Confusion Matrix.

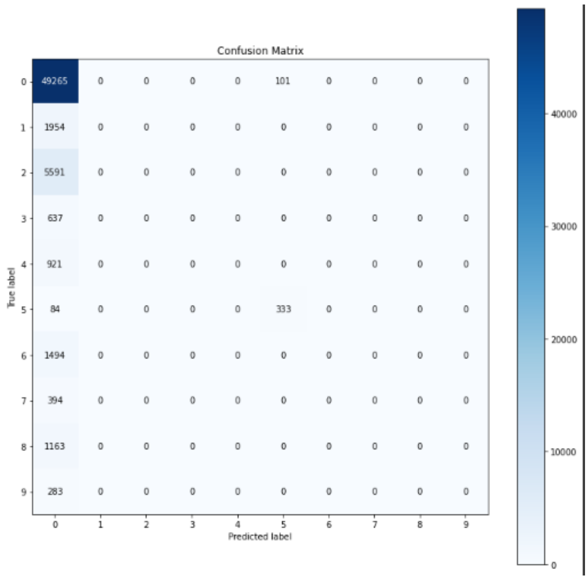

Figure 9: Confusion Matrix of SVM with Linear Kernal

6.2.SVM with RBF Kernel:

Using this algorithm, an accuracy of 79.88 % was achieved. The following figure 10 shows the Confusion Matrix.

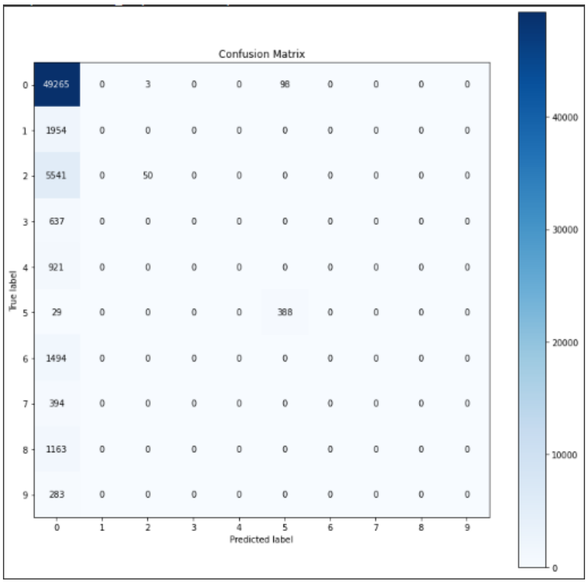

Figure 9: Confusion Matrix of SVM with RBF Kernal

6.2. Multiclass SVM:

Using this algorithm, an accuracy of 49.38 % was achieved. The following figure 11 shows the Confusion Matrix.

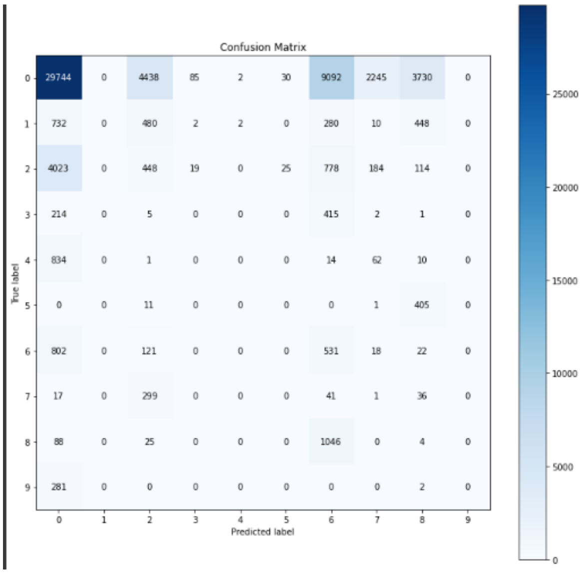

Figure 11: Confusion Matrix of Multiclass SVM

7. Conclusion:

This term paper presented the concept that how Support Vector Machine algorithm which is initially a binary classifier can be used for multiple class problem which is the case of satellite images. It was found that on projecting the initial data onto a higher dimension can help in solving this problem. A thorough accuracy assessment was done to see how the classifier performed with multiple classes present in a hyperspectral satellite image. In the approach followed in this term paper, the data was first pre-processed and then feature extraction was done to extract the most information using the uncorrelated bands and finally SVM classification was carried out using three models: i) SVM with Linear Kernel ii) SVM with Radial Bias Function (RBF) Kernel iii) Multiclass SVM and the accuracy achieved was 79.71%, 79.88%, 49.38% respectively. It can be said that the SVM is a good classification algorithm, and it works well with high dimensional data but when the size of the data increasing, the model takes too much time to train.

References:

[1] J. B. J and J. F. D, "Automatic annotation of satellite images with multi class support vector machine," *Earth Sci. Informatics*, vol. 13, no. 3, pp. 811–819, Sep. 2020, doi: 10.1007/S12145-020-00471-8/FIGURES/5.

[2] "Support Vector Machine for Remote Sensing image classification." https://www.researchgate.net/publication/272493507_Support_Vector_Machine_for_Remote_Sensing_image_classification (accessed Apr. 21, 2022).

[3] G. Anthony, H. Greg, and M. Tshilidzi, "Classification of Images Using Support Vector Machines," Sep. 2007, doi: 10.48550/arxiv.0709.3967.

[4] J. A. Valero Medina and B. E. Alzate Atehortúa, "Comparison of maximum likelihood, support vector machines, and random forest techniques in satellite images classification," *Tecnura*, vol. 23, no. 59, pp. 13–26, Jan. 2019, doi: 10.14483/22487638.14826.

[5] X. Zhou, N.-B. Chang, and S. Li, "Applications of SAR Interferometry in Earth and Environmental Science Research," *Sensors*, vol. 9, no. 3, pp. 1876–1912, 2009, doi: 10.3390/s90301876.

[6] N. Cristianini and E. Ricci, "Support Vector Machines," *Encycl. Algorithms*, pp. 928–932, 2008, doi: 10.1007/978-0-387-30162-4_415.

[7] T. Evgeniou and M. Pontil, "Support vector machines: Theory and applications," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2049 LNAI, pp. 249–257, 2001, doi: 10.1007/3-540-44673-7_12.

[8] "Comparing SVM and GMM on parametric feature-sets | Semantic Scholar." https://www.semanticscholar.org/paper/Comparing-SVM-and-GMM-on-parametric-feature-sets-Mashao/2cdc372a4c48e0cbba932a3ff121e9b6356dfe50 (accessed Apr. 21, 2022).

[9] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," *IEEE Trans. Neural Networks*, vol. 10, no. 5, pp. 1055–1064, 1999, doi: 10.1109/72.788646.