RESEMBLANCE

Team Details:

Team ID - RE1053

Members:

1) Anmol Popli
anmol.ap020@gmail.com
+91-9412648425
En. No: 14115020
Indian Institute of Technology, Roorkee

2) Nitesh Arora
aranitesh101@gmail.com
+91-7060334380
En. No: 14117043
Indian Institute of Technology, Roorkee

3) Pulkit Goyal
pulkitgoyal.work@gmail.com
+91-9012625052
En. No: 14117050
Indian Institute of Technology, Roorkee

INTRODUCTION

The classification of satellite images is part of the process followed for the creation of thematic maps from remotely sensed images. The classification process usually involves three major steps:

a) Segmentation
b) Feature Extraction
c) Classification

Multi-spectral satellite imagery is an economical, precise and appropriate method of obtaining information on land use and land cover since they provide data at regular intervals and is economical when compared to the other traditional methods of ground survey and aerial photography. Classification of multispectral remotely sensed data is investigated with a special focus on uncertainty analysis in the produced land cover maps.

It delivers a great source of data for studying spatial and temporal changeability of the environmental factors. It can be utilized in a number of applications which consists of reconnaissance, making of mapping products for military and civil use, assessment of environmental damage, nursing of land use, radiation level check, urban planning, growth directive, soil test and crop outcome increment. Popular real world applications of Image Classification include identifying, monitoring and analyzing disaster affected regions for effective disaster management; studying and understanding urban encroachment and its consequences and monitoring for changes in hostile territories.

In the current problem assessment; we are provided with satellite imagery for classification, with training datasets for different classes. Dataset consists of four spectral bands- Red, Green, Blue and Infrared, with seven training datasets in each spectral band one for each class (region) to be classified. The task is to train from this given dataset and use it to classify further satellite images into particular regions.

CLASSIFICATION APPROACH

i. Motivation

Two types of approaches are used in satellite image classification:

a) Unsupervised
b) Supervised

Unsupervised classification is an analytical procedure based on clustering. Application of clustering partitions the image data in multispectral space into a number of spectral classes, and then labels all pixels of interest as belonging to one of those spectral classes, although the labels are purely symbolic (e.g. A, B, C, . . . , or class 1, class 2, . . .) and are as yet unrelated to ground cover types. Hopefully the classes will be unimodal; however, if simple unsupervised classification is of interest, this is not essential.
The identification of classes of interest against reference data is often more easily carried out when the spatial distribution of spectrally similar pixels has been established in the image data. This is an advantage of unsupervised classification and the technique is therefore a convenient means by which to generate signatures for spatially elongated classes such as rivers and roads.

In contrast to the a priori use of analyst-provided information in supervised classification, unsupervised classification is a segmentation of the data space in the absence of any information provided by the analyst. Analyst information is used only to attach information class (or ground cover type, or map) labels to the segments established by clustering. Clearly this is an advantage of the approach. However it is a time consuming procedure computationally by comparison to techniques for supervised classification. This can be demonstrated by comparing, for example, multiplication requirements of the iterative clustering algorithm with the maximum likelihood classification decision rule.

A final point that must be taken into account when contemplating unsupervised classification via clustering is that there is no facility for including prior probabilities of class membership. By comparison the decision functions for maximum likelihood classification can be biased by previous knowledge or estimates of class membership. Considering the above mentioned factors we used supervised approach. Suppose some given data points each belong to one of two classes, and the goal is to decide which class a new data point will be in.

The method of maximum likelihood (ML) selects the set of values of the model parameters that maximizes the likelihood function. Intuitively, this maximizes the "agreement" of the selected model with the observed data, and for discrete random variables it indeed maximizes the probability of the observed data under the resulting distribution. Maximum likelihood estimation gives a unified approach to estimation, which is well-defined in the case of the normal distribution and many other problems.

In the case of support vector machines, a data point is viewed as a -dimensional vector (a list of numbers), and we want to know whether we can separate such points with a dimensional hyperplane. This is called a linear classifier. There are many hyperplanes that might classify the data. One reasonable choice as the best hyperplane is the one that represents the largest separation, or margin, between the two classes. So we choose the hyperplane so that the distance from it to the nearest data point on each side is maximized. If such a hyperplane exists, it is known as the maximum-margin hyperplane and the linear classifier it defines is known as a maximum margin classifier; or equivalently, the perceptron of optimal stability.

SVMs perform much better than ML classifier no matter what textural features are used. We think the reasons for that are the number of the samples is very limited and the features may not obey normal distribution, which lead to ML classifier is inefficient and unsuited to our classification.

There are four advantages of SVM: Firstly it has a regularization parameter, which help avoid over-fitting. Secondly it uses the kernel trick, to build in expert knowledge about the problem via engineering the kernel. Thirdly an SVM is defined by a convex optimization problem (no local minima) for which there are efficient methods (e.g. SMO). Lastly, it is an approximation to a bound on the test error rate.

ii. Methodology

Given input training samples {(x1, y1), (x2, y2),……, (xm, ym)}, where xi Є Rd represents a training instance which belongs to a class labelled by yi Є {+ 1,1}. For the not linearly separable training data, the SVM maps the original input space into a high-dimensional feature space via kernel function. The symmetric functions which satisfy the Mercers‟ condition could be served as SVM kernel function. Currently, the popular kernels are used in SVM including: Radial basis function kernel:

$k(xi,xj) = \exp(- \gamma \| xi - xj \| 2)$

Linear kernel: $k(xi,xj) = xi\ T\ xj$
Polynomial kernel: $k(xi,xj) = (xi\ T\ xj + 1)\ d$
Sigmoid kernel: $k(xi,xj) = \tanh [(xi\ T\ xj + b)]$

The separating hyperplane is defined as:

$D(x) = w\ T\ .\ x+b$

Where w is an m-dimension vector, b is a bias term. To obtain the optimal hyperplane, we need to minimize

$$Q (w, b, \xi) = ½ \| w \|2 + C \Sigma \xi i \quad (1)$$

Subject to the constraints

$Yi (w\ T\ xi + b) >= 1 - \xi\ I$ for $I = 1,\ldots,m$

where ξ I are nonnegative slack variables, which measure the degree of misclassification of the datum xi .The constant C is a penalty parameter, which determines the trade-off between maximum classification rate and minimum training error. We call the obtained hyperplane the soft-margin hyperplane, where w is the soft-margin. Equation (1) is a quadratic optimization problem so that it is difficult to solve because of the w. To solve above optimization problem, we can reformulate Equation (1) through a Lagrange function:

$\text{Min max } \{½ \|w\| 2 + C \Sigma \xi I - \Sigma \alpha i [yi (w.xi – b) -1 + \xi I] - \Sigma \beta i \xi I \}$

where αi, βi are the nonnegative Lagrange multipliers. Its dual form is-

$\text{Max } \Sigma \alpha i - ½ \Sigma \alpha i \alpha f\ yi\ yf\ k(xi, xj)$

Subject to the constraints $\Sigma\ y_i\ \alpha_i = 0$, $C >= \alpha_i >= 0$ for $I = 1, \ldots\ldots, m$ where the penalty nonnegative C represents the upper bound here. Finally, we obtain an optimal decision hyperplane

$$D(x) = \Sigma\ \alpha_i\ y_i\ k\ (x_i, x_j) + b$$

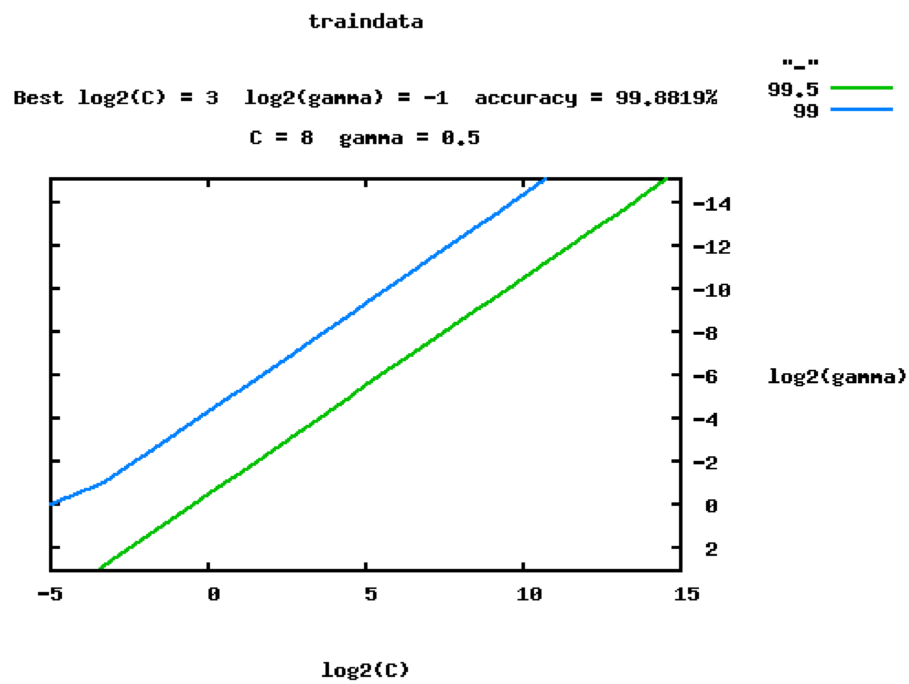
where $k\ (x_i, x_j)$ the kernel function illustrated above. The set S denotes a vector corresponding to the nonzero Lagrange multipliers I, which represents the so-called support vectors (SVs).

iii. Implementation

As part of training, first of all we converted the training windows from the 4 bands to training data stored in one single file. IMG2TRAIN needs to be executed to perform this conversion.

We have used the RBF kernel to linearize our data in a higher dimensional space. This required us to find the optimum value of the parameter gamma of the RBF kernel that linearizes the data to the maximum extent and thus gives the best classification. Also, the cost parameter C was to be optimized for the most violation-free classification of the test data. To tune these parameters, we performed 5-fold cross-validation on the training data for different values of gamma and C. In 5-fold cross validation, the training data is divided into 5 subsets. Then the classifier is trained using 4 subsets and the 5th subset is used as test data to validate the classifier. This is done for all 5 combinations of subsets and the average accuracy of classification is calculated. We performed a grid-search on C and gamma with exponentially growing sequences and arrived at the values C=8 and gamma=0.5 that resulted in the best cross validation accuracy. CROSSVAL is to be executed to perform cross validation.

These optimal parameters were then used to train the classifier with the training data and the classifier model was formed and stored in a file. TRAIN is to be executed to generate the classifier model.



Then we classified the training data using the model generated and computed the accuracy parameters Confusion Matrix, Kappa Coefficient and Overall Accuracy. CLASSIFY_TRAIN is to be executed for classification of training data.

The classifier model can be used to classify test images by executing CLASSIFY_TEST. It reads the test images of 4 bands and converts them to test data stored in a separate file. The test data is classified using the model and classified data with class labels is stored in a file. Also, a colored image is generated that depicts the class labels for all pixels of the input test image.

RESULTS

The training data on being classified by the classifier gives 99.8819% accuracy. It classifies only 1 out of the 847 pixels from the 7 training windows wrongly. This shows that the RBF kernel has linearized the training data well and the cost factor is optimum in that only one vector of the training data has been violated. It can be seen from the output classified image that the test image has been effectively classified into the 7 classes. There are no speckles in the output image and the image is smooth. Regions of same classes have been grouped perfectly with minimal scattering.

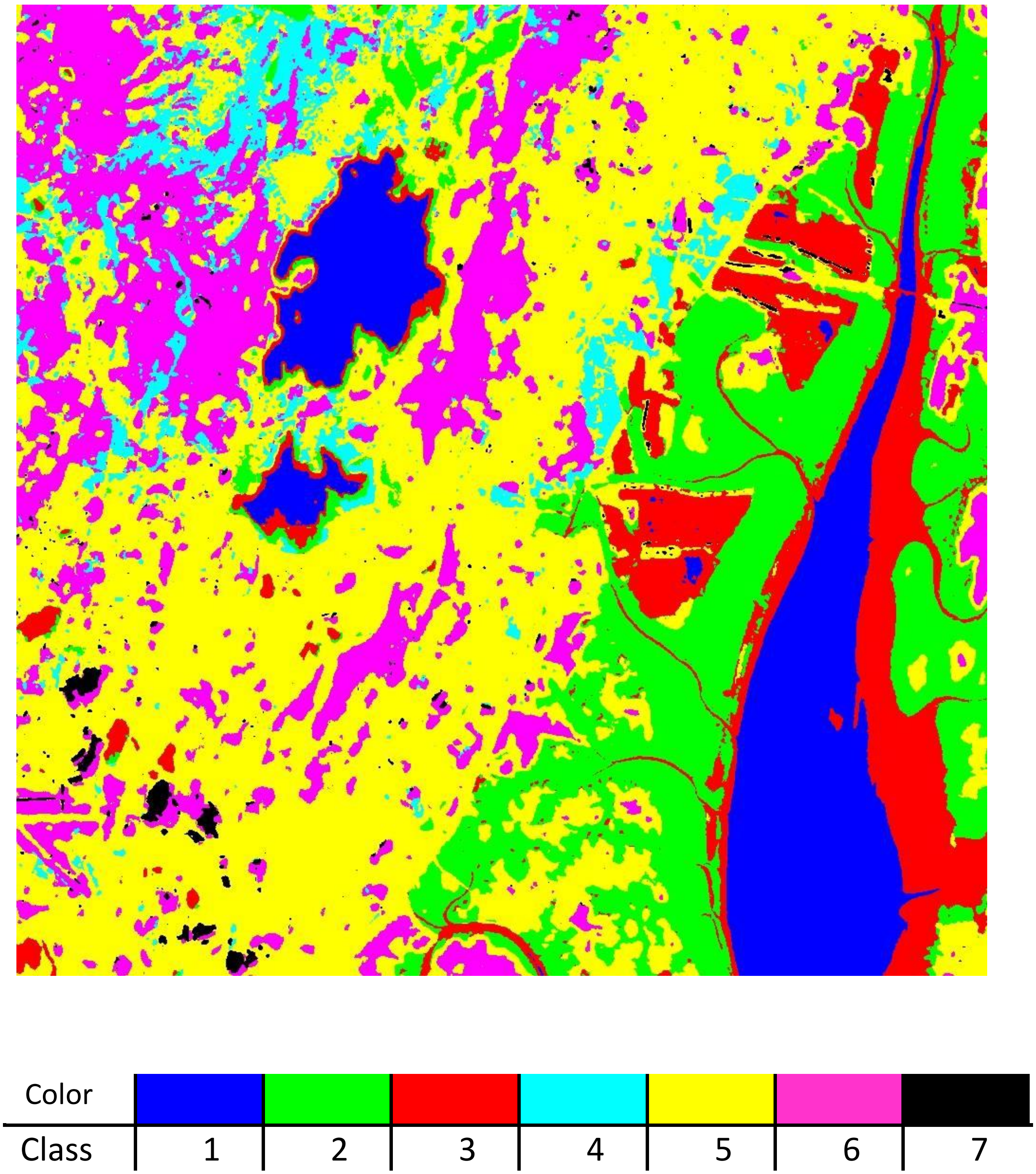


Fig: Image after classification into seven subparts

ACCURACY

i. Confusion Matrix

$$\begin{pmatrix} 121 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 121 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 121 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 121 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 121 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 121 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 120 \end{pmatrix}$$

ii. Kappa Coefficient

Kappa Coefficient = 0.998623

iii. Overall Accuracy

Accuracy = 99.8819 %

CONCLUSION

Our classification method using Support Vector Machine with RBF Kernel is able to linearize data well and also classify it effectively with minimal violation. Thus it gives a high accuracy of 99.8819% in classification of the training data using the classifier model generated. The classified output image is generated by classifying the test images of 4 bands using the trained classifier. Visual analysis of the output image shows that the test image has been classified properly into the 7 classes.

There has been a lot of initiatives to find out methods for reducing the required time for the classification techniques. This has also became essential because of the digital processing and understanding of remote sensing data. Remote sensing and GIS are integral to each other. The development of Remote Sensing is of no use without the development of GIS and vice versa. With remote sensing we not only observe the surface but can also obtain the spatial variability. If the observations are made repeatedly, we can as well obtain the temporal variability. The development and evolution of instruments and research missions of microwave sensing has occurred, certainly, in a quite inhomogeneous and irregular manner. Nevertheless, at the present time, none of potential large-scale satellite missions on earth investigation fails to employ passive and active radio-physical instruments in some configuration. The joint efforts of specialists in various fields geophysicists, computer science and mathematicians; will doubtless be rewarded, and we shall witness some new discoveries in various geophysical disciplines directed at studying the earth.