Effect of Kernel width on classification accuracy: Using SVM for hyperspectral dataset

R. Munsaf Khan, Institute of Space Technology, Pakistan

The advancement in technology has brought into result hyperspectral sensors along with acceptable spatial resolution. At one hand the hyperspectral remote sensing imagery is useful for various applications but on the other hand the processing and interpretation of hyperspectral remote sensing imagery is not a tedious task.

The processing is difficult due to high dimensionality of dataset and atmospheric effects. There are many developed methodologies to reduce the dimensionality of data while preserving maximum information. Research has showed that Support Vector Machines (SVM) generate better results as compared to other dimensionality reduction methods and it also classifies the data along with the reduction.

The major problem associated with SVM is that they were primarily designed to tackle binary problems hence implementation of SVMs onto a hyperspectral dataset can be a laborious task. As the dataset is very complicated hence classes are not linearly separable. An efficient method of non-linearly separation of classes is by use of Kernel function which maps the data into higher dimensional feature space.

There are many types of kernel functions and the result of classification is dependent upon the choice of kernel width. In simpler ways the kernel width is dependent upon the standard deviation we choose for each training class and the choice of standard deviation can greatly affect the classification accuracy even if number of training samples and number of iterations are kept constant.

This research mainly focus on the change in classification accuracy with change in kernel width. A sample data of Salinas is used for this purpose which can be freely downloaded. Results show that small kernel width results in low classification accuracy because in this case the training sample points are over fitted. To choose appropriate size of kernel width a prior information regarding the dataset might be necessary.