UNIT 14

SUPERVISED CLASSIFICATION

Structure

14.1 INTRODUCTION

You have learnt that image classification involves mapping of the digital image data into a finite set of classes that represent surface types in the imagery. It may be used to identify vegetation types, anthropogenic structures, mineral resources, etc. or transient changes in any of these features. Additionally, a classified image that can be considered to a raster layer in a geographic information system. In the previous unit, you have learnt about image classification, typology of image classification and more about unsupervised image classification and its various approaches. In this unit, we will learn about the other generic type of image classification i.e. supervised image classification.

Contributors: Dr. Anupam Anand, Dr. Sourish Chatterjee and Prof. Benidhar Deshmukh

In this unit, we will discuss different image classification methods, signature evaluation, and the guidelines for selecting a classification method.

Expected Learning Outcomes

After studying this unit, you should be able to:

- define supervised image classification;
- describe different supervised image classification algorithms used in remote sensing;
- discuss relative advantages and limitations of commonly used supervised classification algorithms; and
- explain how to evaluate spectral signatures.

14.2 OVERVIEW OF SUPERVISED IMAGE CLASSIFICATION

You are aware that remote sensing images can be recorded from various platforms such as satellites, drones and airplanes. After performing image correction, enhancement and transformation operations on them, the next step is to recognise or classify the pixels in the image into various classes/objects /themes present in the scene based on their spectral signatures or reflectance and remittance properties. This is usually done by a process known as image classification. It may be noted that image classification is used for several societal applications namely, land use/land cover analysis, agriculture, urban planning, natural resource management, surveillance, object detection, updating geographic maps and disaster mitigation.

Image classification can be defined as a process of assigning land cover classes/themes to pixels in an image (Lillesand and Keifer, 1994). Some of the classes comprise built-up area, urban, forest, grassland, agriculture, water, shadow, rocky areas, bare soil and cloud. Image classification usually represents object of the analysis and generates a map-like image in the form of final product/output. It is an important tool for studying digital images. There are several image classification methods available namely, supervised, unsupervised, per-pixel, object-based, hard, soft, parametric, non-parametric, etc.

14.2.1 Classification Scheme

When we go for image classification, the first thing that is important to consider is the classification scheme. It is important because it determines the level of details that is going to be there in the classified output. You have read about LULC classification scheme in Block-3 of MGY-101 and also in MGY-102. You will read about the LULC classification scheme in some more detail in the Unit 5 of the course MGY-007. LULC classification scheme contains taxonomically correct definition of class information that is organised according to a logical criterion. If a class satisfies certain criteria, it will be classified to that class. Defining the criteria helps the analyst, irrespective of the user, and therefore helps in maintaining consistency in classification. A good classification scheme

should avoid overlapping of classes through its definition. For example, there should be a crisp definition to classify an area as agriculture. A scheme designed for classification needs to be exhaustive, such that all features in an area are correctly classified into class, while also taking precaution of excluding the classes that do not occur in a region. For example, an inland area you need not have a class called 'inter-tidal region', which is available only in coastal areas, signifying the land exposed during high-tide and low-tide, but are extremely important from the point of view of biodiversity and conservation of aquatic species.

So, you need to choose a classification scheme based on the theme you are interested in and also the level of classification. It will also determine the choice of data that is suitable for mapping the classes of your interest.

Let us now recall the stages of classification.

14.2.2 Stages

You have read in the previous unit i.e. Unit 13 that as the name implies, unsupervised classification is carried out without interpretive guidance from an analyst. An algorithm automatically organises pixels having similar spectral properties into groups that become the basis for different classes. This is entirely based on the statistics of the image data distribution, and is often called clustering. The process is automatically optimised according to cluster statistics without the use of any knowledge-based control (i.e. ground referenced data). The method is, therefore, objective and entirely data driven. It is particularly suited to images of targets or areas where there is no ground knowledge. Even for a well-mapped area, unsupervised classification may reveal some spectral features which were not apparent beforehand.

Supervised classification, as the name implies, requires human guidance. An analyst selects groups of contiguous pixels from the input image known as training areas that defines DN values in each channel for each class. A classification algorithm computes certain properties (i.e. data attributes) of the training pixels, e.g. mean DN for each channel. Then, DN values of each pixel in the image are compared with the attributes of the training set. This is based on the statistics of training areas representing different ground selected subjectively by users on the basis of their own knowledge or experience. Classification is controlled by users' knowledge but, on the other hand, is constrained and may even be biased by their subjective view. Classification can, therefore, be misguided by inappropriate or inaccurate training area information and/or incomplete user knowledge. A standard approach for carrying out supervised image classification is given in Fig. 14.1.

14.3 SELECTION OF TRAINING SITES AND SIGNATURE EVALUATION

As you have read earlier, evaluation of signatures is an important step in classification, which is carried out before decision making stage. In this stage, signatures of different classes obtained through training sites from image are checked for their representativeness of class they attempt to describe and also

to ensure their uniqueness from other classes. However, let us first review the concept of spectral signature and then about the ways of signature evaluation.

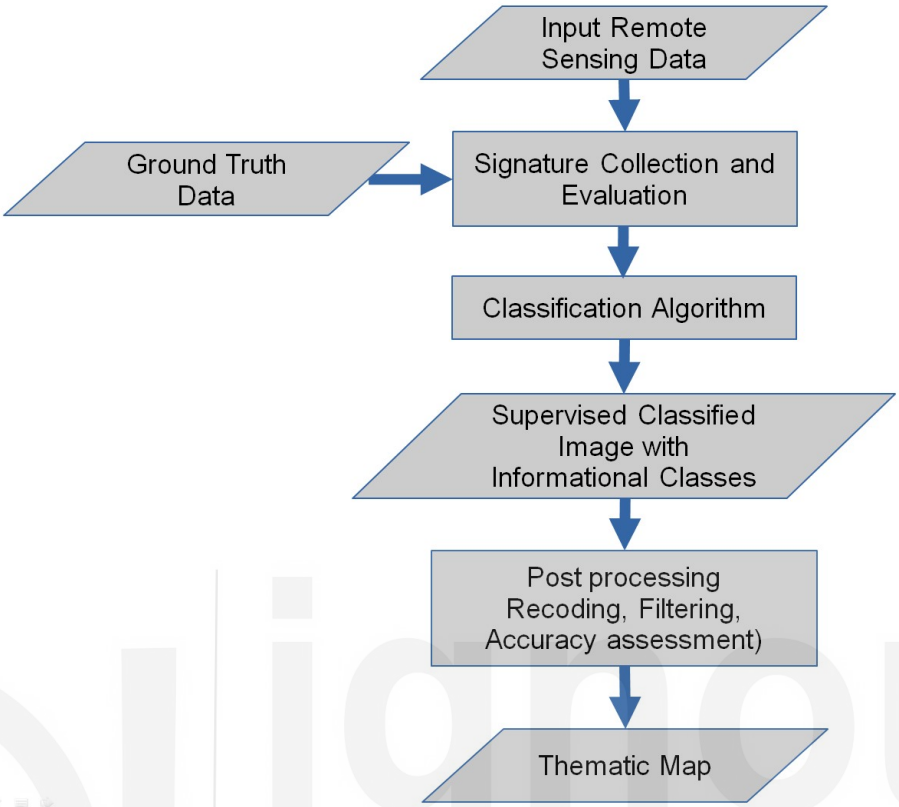

Fig. 14.1: Typical stages in the process of supervised image classification.

14.3.1 Spectral Signature

The wavelength of any given material determines the amount of solar radiation it reflects, absorbs, transmits, or emits. So, when the amount of solar radiation reflected, absorbed, transmitted, or emitted (usually measured in intensity, as a percent of maximum) by the material is plotted over a range of wavelengths, the connected points produce a curve called the material's spectral signature.

The percent reflectance values of similar objects at a selected wavelength will be similar while it will vary for different objects or landscape features. These values can be plotted in a graph and compared. Such plots are called spectral response curves or spectral signatures. Spectral signatures of like features have similar shapes, for example, concrete will have similar spectral signatures while the spectral signatures of grass and concrete will vary. Differences among spectral signatures are used to classify remotely sensed images into classes of landscape features.

Greater details of recorded spectral information allow for greater information to be extracted from spectral signatures. This important property of matter makes it possible to identify different substances or classes and also to separate them by their individual spectral signatures (Fig. 14.2).

For example, at some wavelengths, soil reflects more energy (absorbs less) than green vegetation but at other wavelengths it absorbs more (like clayey soil) than does the vegetation. These differences in reflectance from various kinds of surface materials make it possible to differentiate them from one another.

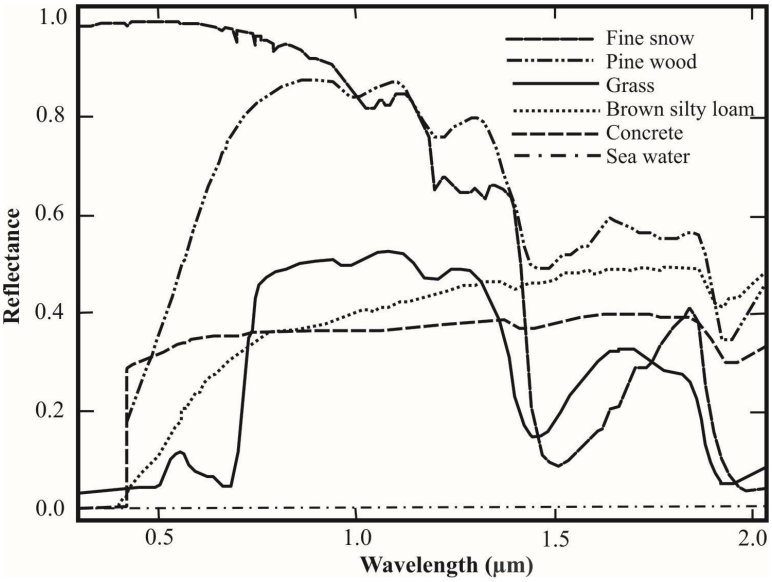

Fig. 14.2: Spectral signature of various features.

Spectral response curves of some of the materials are discussed here.

Vegetation contains water, cellulose (tissues and fibres), constituent of wood, lignin (non-carbohydrate nitrogen), chlorophyll (green pigments) and anthocyanin (water-soluble pigments). Depending on how 'active' (i.e. kinds of chlorophyll) a green vegetation is, the combination of transmittance, absorbance and reflectance vary in different bands of the spectrum. Here is a general example of a reflectance plot for a vegetation type, with the dominating factor influencing each interval of the curve so indicated; note downturns of the curve that result from selective absorption (Fig. 14.2).

Chlorophyll strongly absorbs radiation in the red and blue wavelengths but reflects green wavelengths. Leaves appear "greenest" to us in the summer, when chlorophyll content is at its maximum. In autumn, there is less chlorophyll in the leaves, so there is less absorption and proportionately more reflection of the red wavelengths, making leaves appear red or yellow (yellow is a combination of red and green wavelengths). The internal structure of healthy leaves acts as excellent diffuse reflectors of NIR wavelengths. If our eyes were sensitive to NIR, trees would appear extremely bright to us at these wavelengths.

Overall, factors such as leaf damage, sun and shade, leaf water content; leaf air spaces and salinity and nutrient levels can affect spectral response of the leaf.

The spectral response of vegetation canopies is a little different from that of leaves. Transmittance of leaves, amount and arrangement of leaves, structural characteristics such as stalks, trunks, limbs; background (soil, leaf litter, etc.); solar zenith angle; viewing angle and azimuth angle influence the spectral response.

Soil tends to have reflection properties that increase approximately monotonically with wavelength. They tend to have high reflectance in all bands. This is dependent on factors such as colour, constituents and especially the moisture content. As described above, water is a relatively strong absorber of all wavelengths, particularly those longer than the red part of visible spectrum

Contributors: Dr. Anupam Anand, Dr. Sourish Chatterjee and Prof. Benidhar Deshmukh

(Fig. 14.3). Therefore, as soil moisture content increases, the overall reflectance of that soil tends to decrease. Soils rich in iron oxide reflect proportionally more of the red than other visible wavelengths and therefore appear red (rust colour) to the human eye. A sandy soil on the other hand tends to appear bright white in imagery because visible wavelengths are more or less equally reflected; when slightly less blue wavelengths are reflected this results in a yellow colour. In a nutshell, spectral response curves of soil and rocks are influenced by soil colour, mineral content, inter-molecular vibration of the molecules, organic matter (influences soil colour and moisture), particle size, reflectance and thermal diffusivity and moisture (Fig. 14.3).

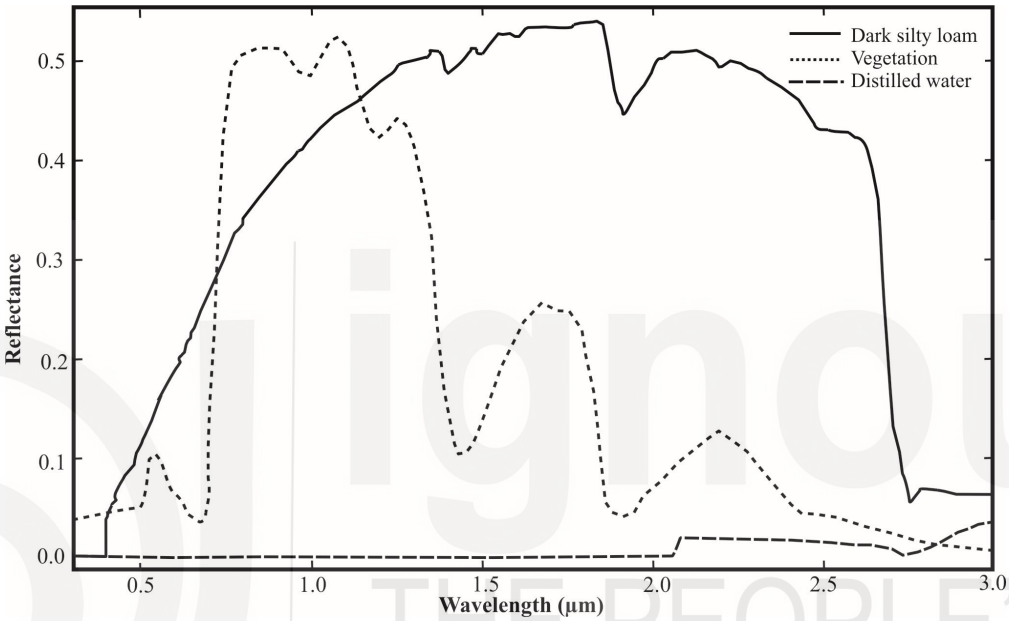

Fig. 14.3: Generalised spectral signatures for some of the features.

Water absorbs much longer wavelength at visible and NIR radiation than shorter visible wavelengths. Thus, water typically looks blue or blue-green due to stronger reflectance at these shorter wavelengths and darker if viewed at red or NIR wavelengths. If there is suspended sediment present in the upper layers of a water body, then this will allow better reflectivity and a brighter appearance of the water. The apparent colour of the water will show a slight shift to longer wavelengths. Suspended sediment can be easily confused with shallow (but clear) water, since these two phenomena appear very similar. Chlorophyll in algae absorbs more of the blue wavelengths and reflects green, making water appear greener in colour when algae are present. The topography of the water surface such as rough, smooth or floating materials can also lead to complications for water related interpretation due to potential problems of specular reflection and other influences on colour and brightness.

Spectral signatures, however, are not always “pure” which means the sensor might record some signatures that may be emitted by surrounding objects. “Pure” spectral signature for individual materials or classes can be determined best under laboratory conditions, where the sensor is placed very close to the target (Fig. 14.4). There is no interference in a closed and controlled environment such as a laboratory. Agencies such as ISRO, US Department of Agriculture and several universities maintain large repositories of spectral

signatures. Moreover, many image analysis tools have such spectral libraries bundled with the software.

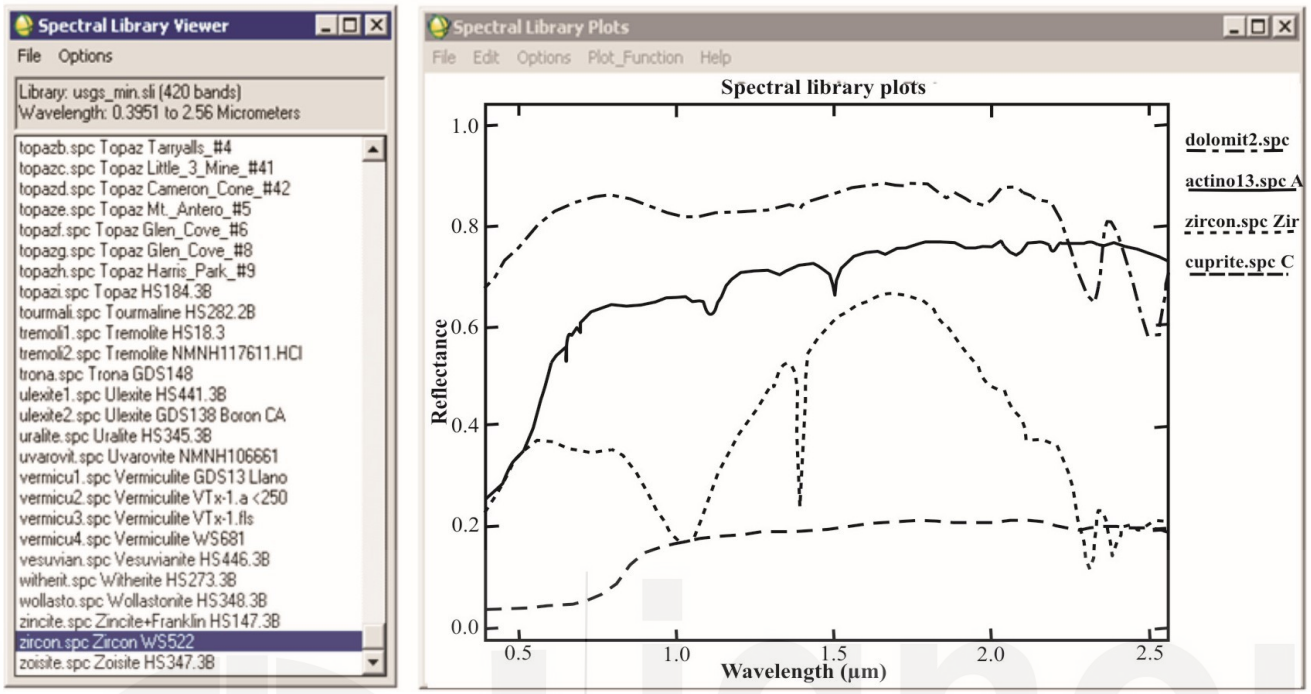

Fig. 14.4: Spectral curves from a spectral library. (Source: Spectral library of US Geological Survey)

You have leant that spectral signature is the unique pattern of reflectance or radiance values of an object or surface across various wavelengths of the electromagnetic spectrum. Understanding spectral signature is important because they are used to differentiate between different classes or materials in remote sensing images. Accurate signatures are crucial for distinguishing between similar classes and improving classification accuracy.

Let us now study about training site selection.

14.3.2 Training Site Selection

After selection of suitable classification scheme and the data for classification, the supervised image classification process consists of three important stages: training, signature evaluation and decision making as you have read earlier (Fig. 14.5).

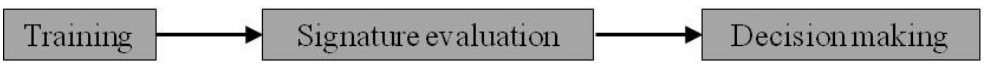

Fig. 14.5: Typical stages in the process of supervised classification.

Training is the process of generating spectral signature of each class. For example, a forest class may be defined by minimum and maximum pixel values in different image bands, thus defining a spectral envelope for it.

This simple statistical description of the spectral envelope is known as signature. Training can be carried out either by an image analyst with guidance from his experience or knowledge (i.e. supervised training) or by some statistical clustering techniques requiring little input from image analysts (i.e. unsupervised training).

Selecting training sites for supervised image classification involves strategically choosing specific regions or samples from the image data to build a model that can accurately classify various classes within the image. Prior to the selection of training sites, you need to clearly know the classes or categories that you want to classify. For example, in land cover classification, classes could be forest, water, urban areas, etc. You also need to have understanding of the characteristics of each class in terms of their visual image interpretation so that you can identify them in the image. You also need to perform appropriate preprocessing steps such as radiometric correction, atmospheric correction, and normalisation to ensure the data is suitable for classification. Domain knowledge is important to correctly identify the representative areas (for the features of interest) based on visual inspection in the image and select initial training sites. The sites are manually selected which involves drawing polygons or selecting points that represent different classes.

The training data selection is strongly governed by the theory of sampling that is a part of the field of statistics. Small classes or highly homogeneous classes can have a small training dataset while a large or heterogenous class requires a larger training dataset. The training dataset is expected to capture all the intra-class variation given the pixel vectors within each class. If you take very less number of training sites then it may be difficult to obtain a spectral signature which truly represents that class. It must be ensured that the location of each training sample and its thematic class are correctly recorded, else the classification result can be erroneous.

We have read that signature Evaluation is the checking of spectral signatures for their representativeness of the class they attempt to describe and also to ensure as small spectral overlap between signatures of different classes as possible.

Decision making is the process of assigning all the image pixels into thematic classes using evaluated signatures. It is achieved using algorithms, which are known as decision rules. The decision rules set certain criteria. When signature of a candidate pixel passes the criteria set for a particular class, it is assigned to that class. Pixels failing to satisfy criteria of any of the classes remain unclassified. We shall discuss in detail the decision rules in the next two sections.

You may note here that the training and signature evaluation steps are essential steps in supervised classification whereas in the case of unsupervised classification, training and signature evaluation is not involved and the focus is on assigning a thematic class to the classes generated by the computer through minimum input by the human analyst.

14.3.3 Ways of Signature Evaluation

One of the most common techniques for feature identification is spectral evaluation. Most of the image analysis software provides an interface to plot spectral signature. Fig. 14.6 shows an example of how a spectral image is plotted using an image analysis tool. With knowledge about the spectral profile for a given feature, we can go back and change band combinations to make that feature show up more clearly on the image.

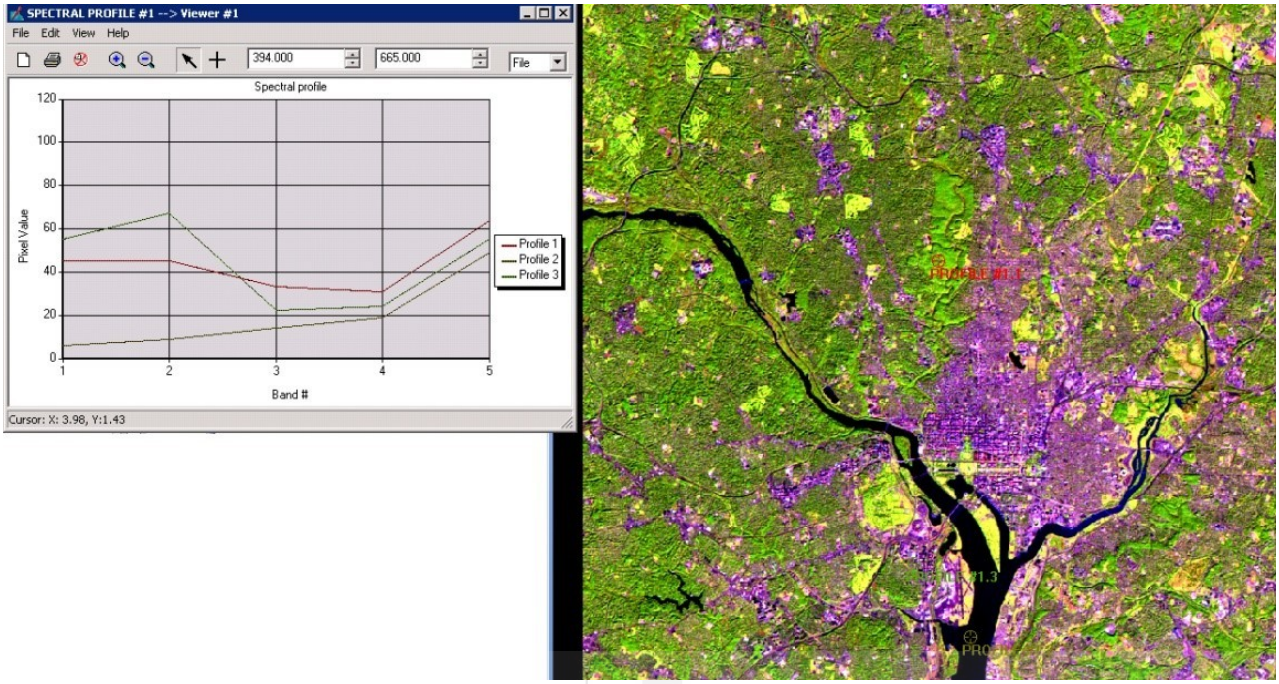

Fig. 14.6: Spectral plots from a satellite image.

Signature evaluation is a critical step in supervised image classification, as it ensures that the selected training samples accurately represent land cover classes. Evaluating spectral signatures helps refine class definitions, minimise misclassification, and improve classification accuracy. Signature evaluation involves assessing the statistical separability, spectral consistency, and classification performance of training samples before applying any classification algorithms.

Signature evaluation is performed using qualitative and quantitative approaches that include visual analysis, statistical measures, and classification accuracy prediction. Spectral signatures are evaluated in the following ways:

- Spectral signature analysis
- Statistical separability analysis
- Feature space visualisation
- Signature refinement and optimisation
- Accuracy prediction and preliminary classification

- **Spectral signature analysis:** classification is performed on the pixels within the training samples for each class and is compared with classes as recorded in the field data on those locations. Ideally, all pixels in a training sample should classify correctly. However, you can expect high percentages of correctly classified pixels if the signatures taken are appropriate. You can examine spectral response patterns of different land use land cover classes to ensure uniqueness and consistency. To inspect signature of training samples, spectral profiles of training samples are plotted across multiple bands. It helps assess whether different land use and land cover types have distinct spectral characteristics. If spectral profiles of different classes overlap significantly, they may be spectrally indistinguishable, leading to classification errors. In such cases, additional features (e.g., texture, indices, or ancillary data) can be incorporated.

- **Statistical separability analysis:** statistical measures that you have studied in Block-1 of this course such as mean and standard deviation of each signature are used to plot ellipse diagrams in two or more dimensions. The plotting allows the analyst to identify similar signatures and hence the classes which are likely to suffer most from misclassification. If the amount of overlap between a pair of signatures is large then those classes are not separable using that image data. Large standard deviations indicate high internal variability within a class, which may lead to misclassification. Scatter plot is used to visualise the distribution of different land cover classes in feature space (e.g., Red vs. NIR bands). Histograms can also be used to show pixel value distributions to detect overlap between classes. If histograms for two classes overlap significantly, it can be said that they may not be well-separated. It is important to ensure that there is high separability between signatures from different types of training samples and low separability among signatures from the training samples of a particular class. Distance based separability measures quantify the spectral distinctiveness between two classes using distance metrics such as by measuring spectral distance, i.e. separability by computing Euclidean distance, Bhattacharyya distance, Jeffries-Matusita distance or transformed divergence. You can find mathematics behind computation of these in a book by Swain and Davis (1978). However, let us briefly learn about them.

  Euclidean distance measures the straight-line distance between class means in spectral space. Larger distances indicate better separability. The Euclidean Distance, D between two points P_1 (x_1, y_1) and P_2 (x_2, y_2) in a two-dimensional space is given by:

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

  For n-dimensional feature space (e.g., multiple spectral bands in remote sensing), Euclidean Distance between two feature vectors $X = (x_1, x_2, ..., x_n)$ and $Y = (y_1, y_2, ..., y_n)$ is:

$$D(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

  where, x_i and y_i represent spectral values in the i-th band for two different pixels, classes, or training samples. Suppose we want to classify a pixel into either Water or Vegetation based on NIR and Red bands. And the mean spectral values for Water is (NIR = 50, Red = 30) and for Vegetation is (NIR = 150, Red = 60). If we calculate Euclidean Distance for a pixel with values (NIR = 100, Red = 45), to water it would be:

  $D = (100 - 50)^2 + (45 - 30)^2 = 2500 + 225 = 2725 \approx 52.2$

  And to Vegetation it would be:

  $D = (100 - 150)^2 + (45 - 60)^2 = 2500 + 225 = 2725 \approx 52.2$

  In this case, since distances are equal, additional spectral bands or features are needed for better classification. It is easy to compute, however, it may

not be suitable for highly mixed or overlapping classes, requiring additional statistical methods for improved accuracy.

Bhattacharyya distance measures the statistical overlap between two class distributions. It helps evaluate class separability by comparing the probability distributions of spectral values across multiple bands. A higher Bhattacharyya Distance indicates better class separability, while a lower value suggests significant overlap, making classification more challenging. The values closer to 2 indicate better separability, while values near 0 suggest high overlap. The Bhattacharyya Distance between two probability distributions P(x) and Q(x) is defined as:

$$D_B = -\ln\left(\sum_x \sqrt{P(x)Q(x)}\right)$$

For multi-dimensional Gaussian distributions, commonly used in remote sensing applications, the Bhattacharyya Distance between two classes i and j is given by:

$$D_B = \frac{1}{8}(\mu_i - \mu_j)^T \Sigma^{-1}(\mu_i - \mu_j) + \frac{1}{2}\ln\left(\frac{|\Sigma|}{\sqrt{|\Sigma_i||\Sigma_j|}}\right)$$

where:

μ_i, μ_j are the mean vectors of the two classes.

Σ_i, Σ_j are the covariance matrices of the two classes.

Σ is the average covariance matrix:

$$\Sigma = \frac{1}{2}(\Sigma_i + \Sigma_j)$$

$|\Sigma|$ represents the determinant of the matrix Σ.

If DB = 0, the two classes are completely identical meaning there is no separability. If DB > 2, the two classes are highly separable. And if 0 < DB < 20, partial class overlap exists, requiring additional spectral or contextual information for better classification.

The metric is advantageous as unlike Euclidean Distance, it considers the distribution and spread of spectral values. It is more reliable than simple distance metrics as it provides a probabilistic approach to class separability. It helps identify the most effective bands in hyperspectral and multispectral classification. However, it has certain limitations such as: it is computationally intensive, sensitive to small sample sizes, and is less effective for non-Gaussian distributions.

Jeffries-Matusita (JM) distance is similar to Bhattacharyya Distance but bounded between 0 (high overlap) and 2 (perfect separability). It is recommended for evaluating training sample quality. The JM Distance is particularly useful in training data evaluation, feature selection, and supervised classification to determine whether different classes can be effectively distinguished.

The JM Distance between two probability distributions *P(x)* and *Q(x)* is given by:

$$JM = 2\left(1 - e^{-B}\right)$$

where, *B* is the Bhattacharyya Distance, calculated as discussed earlier.

Values of JM distance ranges between 0 and 2. If JM = 0 it means that there is no separability meaning that the classes are identical. If JM = 2, it means that there is perfect separability (classes are completely distinct). If JM > 1.9, it means that there is good separability, meaning that it is ideal for classification. If JM < 1, it means that there is high class overlap, that may lead to misclassification.

The Jeffries-Matusita (JM) distance is a powerful statistical tool for evaluating spectral separability in remote sensing image classification. It helps in assessing training data quality, selecting optimal spectral bands, and improving classification accuracy. It is advantageous as ensures that training data are well-separated before classification. It is more reliable than simple distance metrics as it incorporates both mean and variance of spectral distributions. It is useful for feature selection for hyperspectral data as it helps reduce dimensionality while retaining meaningful information. However, it has certain limitations such as it is computationally intensive, assumes Gaussian distribution and is sensitive to small sample sizes.

Transformed divergence measures how different two probability distributions are. It is a statistical separability measure used to evaluate the distinction between spectral signatures of different land cover classes. It is derived from the Divergence measure, which quantifies the statistical difference between two probability distributions. The Transformed Divergence applies a nonlinear transformation to improve its interpretation and scaling, making it widely used for training sample evaluation and feature selection in supervised classification.

The Divergence (D) between two classes **i** and **j** with probability distributions P(x) and Q(x) is given by:

$$D = \frac{1}{2}\mathrm{Tr}\left(\Sigma_i^{-1}\Sigma_j + \Sigma_j^{-1}\Sigma_i - 2I\right) + \frac{1}{2}(\mu_i - \mu_j)^T\left(\frac{\Sigma_i^{-1} + \Sigma_j^{-1}}{2}\right)(\mu_i - \mu_j)$$

where:

μ_i, μ_j are the mean vectors of the two classes.

Σ_i, Σ_j are the covariance matrices of the two classes.

Tr() denotes the trace of a matrix.

The Transformed Divergence (TD) applies a nonlinear transformation to D to ensure that values are more meaningful for class separability evaluation:

$$TD = 200 \times \left(1 - e^{-\frac{D}{8}}\right)$$

Values of TD range from 0 to 200**,** providing an easily interpretable scale. Higher divergence values indicate better class distinction. If TD=200, it

represents perfect class separability which is ideal for classification. If TD>150, it represents good separability meaning reliable classification. If 100<TD<150, it represents moderate separability meaning that it may require additional features and if TD<100, it represents poor separability meaning significant spectral overlap exists between the classes.

A comparison of the four distance metrics is given in Table 14.1.

Table 14.1: Comparison of the commonly used supervised classification algorithms:

Criterion	Euclidean Distance	Bhattacharyya Distance	Jeffries-Matusita (JM) Distance	Transformed Divergence (TD)
Description	Measures the straight-line distance between two points in spectral space	Measures statistical separability based on probability distributions	Derived from Bhattacharyya Distance, scaled between 0 and 2	A nonlinear transformation of Divergence to a 0–200 scale
Type of measure	Geometric distance	Probabilistic distance	Probabilistic distance	Probabilistic distance
Accounts for variability (Covariance)?	No	Yes	Yes	Yes
Range of values	0 to ∞	0 to ∞	0 to 2	0 to 200
Interpretability	Simple and intuitive	More complex, but indicates spectral separability	Intuitive (higher values mean better separation)	Easy to interpret (scaled for intuitive understanding)
Computational complexity	Low (Fast)	High (Computationally intensive)	High (Requires Bhattacharyya Distance)	High (Matrix operations required)
Sensitivity to data distribution	No (does not account for class distribution)	Yes (considers mean and variance)	Yes (considers mean and variance)	Yes (considers mean and variance)

- **Feature space visualisation:** It is used to visually assess how well classes are distributed in multi-dimensional feature space. For the purpose, Principal Component Analysis (PCA), box plots and density plots are used. PCA reduces high-dimensional spectral data into principal components and helps to visualise clusters of training samples to identify spectral confusion. Box plot shows the range, median, and variability of spectral values for different classes. Overlapping boxes indicate poor class distinction. Density plot displays how pixel values are distributed within each class and is useful for detecting outliers and mixed pixels.
- **Signature refinement and optimisation:** It improves class separability by modifying training samples. There are three approaches used for the purpose, viz. outlier detection and removal, merging and splitting classes and use of ancillary data. In the process of outlier detection and removal outliers that distort statistical properties of training signatures identified and removed through Mahalanobis Distance, Box Plot Analysis, or K-Means Clustering. In the process of merging and splitting classes, if two classes are highly confused, they can be merged into one broader category. And, if

a class exhibits high variability, it can be split into sub-classes for better representation. In the third approach, ancillary data such as elevation, slope, or texture features that can help separate spectrally similar classes are used e.g. DEM data can differentiate forest types in mountainous areas.

- **Accuracy prediction and preliminary classification**: It is used to assess how well training samples perform before full classification. One of the approaches is cross-validation of training samples in which training samples are divided into training and testing subsets and preliminary classification is performed and accuracy assessed. In the second approach, training samples are classified using a classifier and results are compared against known labels to detect poorly classified samples. Spectral Angle Mapper (SAM) is also used for assessing class consistency. It measures the spectral similarity between training samples and reference data. It is useful for hyperspectral data where spectral shape is more important than absolute values.

You should note that some of the training samples whose signatures have negative effect on the classification outcome need to be either renamed or merged or deleted.

14.3.4 Selection of Optimum Number of Bands

Since, remote sensing images contain multiple spectral bands, each capturing different spectral properties of the features present in the study area, choosing the optimum number of bands for image classification is crucial for improving the accuracy and efficiency of the classification process. Selecting the right number of bands involves balancing between capturing sufficient information for accurate classification and avoiding redundancy and computational inefficiency.

Selection of the optimum number of bands is determined by the following factors:

Context and objectives: You may select the bands suitable for mapping the features based on the specific objectives of the classification task such as mapping land use land cover classes, vegetation, waterbodies, or minerals/rocks. It would be guided by the characteristics of the data being used.

Information contained in the bands: Since, each band provides unique information, and some may be more relevant to the classification task than others depending upon their characteristics. Further, the correlating bands may be avoided and highly correlated bands might provide redundant information. Correlation of the spectral bands can be calculated using correlation analysis or Principal Component Analysis (PCA). PCA is used to transform the data into principal components and select the number of components that explain a significant portion of the variance. This helps in reducing dimensionality while retaining important information.

Computational efficiency: Higher number of bands means more computational resources and processing time. You can choose a number of bands that balances classification accuracy with computational efficiency and also the storage requirements and manageability of the dataset with different numbers of bands.

Class separability: Band combinations that enhance class separability should be prioritised. Effect of inclusion or exclusion of certain bands on separability between classes may be assessed and accordingly the optimum number of bands may be selected.

Choosing the optimum number of bands for image classification involves evaluating the relevance and redundancy of each spectral band, applying feature selection techniques, and balancing classification performance with computational efficiency. By systematically analysing band correlations, applying dimensionality reduction methods, and testing various band combinations, you can identify the most effective set of bands for classification.

Let us spend 5 minutes to check your progress.

SAQ I

a) What is supervised image classification?

b) What is a training site?

c) What do you understand by signature evaluation?

d) What are the ways of signature evaluation?

In the following sections, we will discuss about approaches of supervised classifications.

14.4 APPROACHES OF SUPERVISED CLASSIFICATION

As you have read earlier, supervised classification, as the name implies, requires human guidance. An analyst selects groups of contiguous pixels from the input image known as training areas that defines DN values in each channel for each class. A classification algorithm computes certain properties (data attributes) of the training pixels, for example, mean DN for each channel (Fig. 14.7). Then, DN values of each pixel in the image are compared with the attributes of the training set.

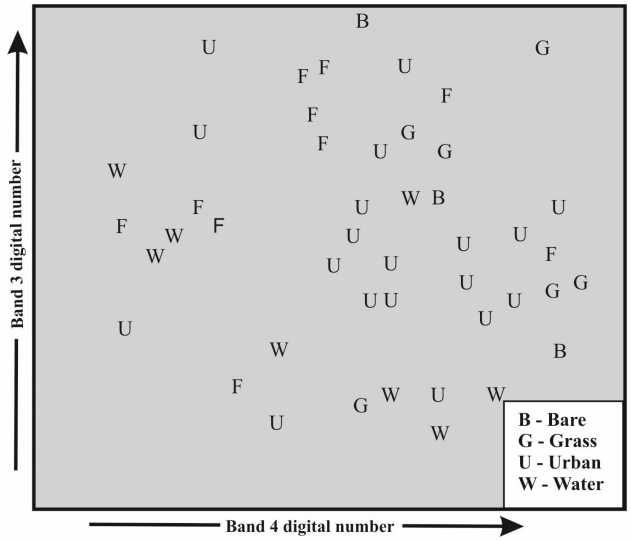

Fig. 14.7: Using supervised classification, pixels are classified into different categories.

This is based on the statistics of training areas representing different ground objects (Fig. 14.8) selected subjectively by users on the basis of their own knowledge or experience. Classification is controlled by users' knowledge but, on the other hand, is constrained and may even be biased by their subjective view. Classification can, therefore, be misguided by inappropriate or inaccurate training area information and/or incomplete user knowledge.

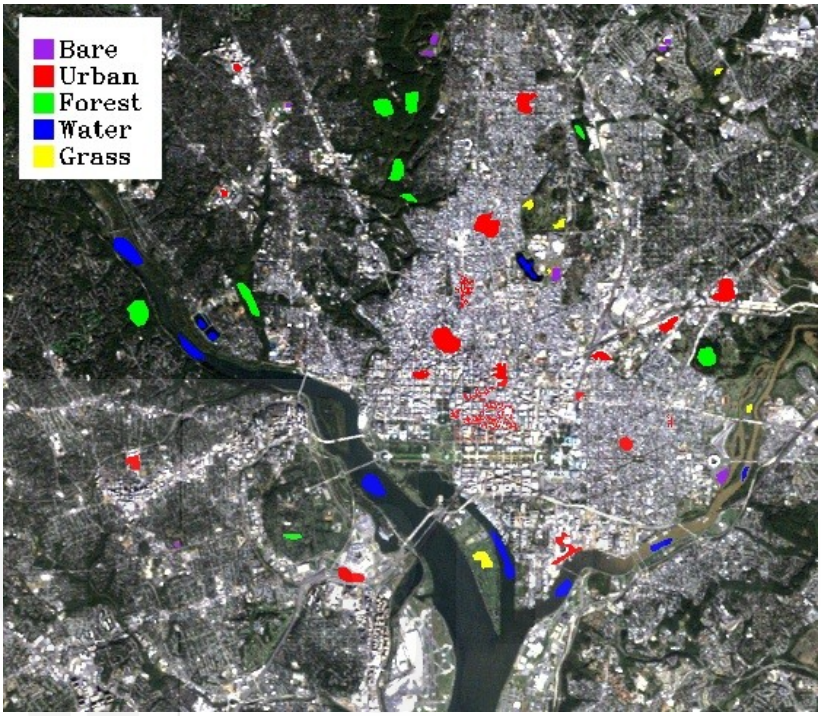

Fig. 14.8: Locations of the training data collected for supervised classification.

In the following subsections, we will discuss some of the commonly used algorithms such as minimum distance to means, parallelepiped and maximum likelihood algorithms, etc.

14.4.1 Minimum Distance to Means Algorithm

The Minimum Distance to Means (MDM) algorithm is a straightforward and commonly used supervised classification method in remote sensing and pattern recognition. It is based on the assumption that each class in the image can be represented by a central point (a mean vector) in the feature space. **Feature space** is a multi-dimensional space where each dimension represents a different attribute or feature extracted from the original data. Here in this case, it represents a scatterplot in which position of each pixel is determined by the spectral values of the two selected bands. This central point is usually calculated from the training samples of the class. The algorithm assigns each pixel or data point to the class whose mean (or centroid) is closest. This method leverages the concept of Euclidean distance to determine the nearest class centroid.

Classification using this classifier is carried out in the following major steps:

Step 1: Training Phase:

- Data (i.e. Training Samples) collection: Collect representative training samples for each class. These samples have known class labels and serve as the basis for calculating class centroids.

- Calculate Mean Vectors for each class: Compute the mean vector (centroid) for each class by averaging the pixel values (or feature values) of the training sets belonging to that class. For example, if you have a multi-band image, each class would have a mean vector representing the average spectral values across all bands.

Mathematically, the mean for class C is computed as:

$$\mu_C = \frac{1}{N_C} \sum_{i=1}^{N_C} x_i$$

where,
μ_C is the mean vector for class C,
x_i are the feature vectors of the training samples for class C, and
N_C is the number of training samples for class C.

Step 2: **Classification Phase**:

- Compute distance: For each pixel (or data point) in the image, calculate the distance between the pixel and the mean vector of each class.
- Assign class: Assign the pixel to the closest i.e. class with the smallest distance (generally, the Euclidean distance). Mathematically, it is represented as

$$d(x, \mu_C) = \sqrt{\sum_{i=1}^{n} (x_i - \mu_{C,i})^2}$$

where,
x is the feature vector of the pixel
μ_C is the mean vector for class *C*, and
n is the number of features (e.g., bands).

Step 3: **Output:**

- Each pixel is assigned a class label based on the minimum distance to the class means.

Let us now see the advantages and limitations of this classifier.

Advantage

- Simplicity: The MDM algorithm is easy to understand, implement, and interpret.
- Computational efficiency: It is computationally efficient, requiring only the calculation of mean vectors and Euclidean distances, making it suitable for large datasets.
- Quick training: The training phase is straightforward and quick, involving only the computation of mean vectors from the training data.
- Interpretability: The results are easy to interpret since they are based on the proximity of a pixel to predefined class centroids.

Limitation

- Assumes linearity and class normality: It assumes that the classes are linearly separable and each class can be well-represented by a single centro-

id, which may not always be the case especially for classes with high variability.

- Sensitivity to outliers: The algorithm can be sensitive to outliers in the training data, which can skew the mean vectors and affect classification accuracy.
- Equal variance assumption: It assumes that all classes have equal variance, which might not hold true for all datasets, potentially leading to misclassification.
- Limited complexity handling: The MDM algorithm may not perform well in complex scenarios where class distributions are not well-defined or significantly overlap.

The Minimum Distance to Means algorithm is a straightforward supervised classification technique which is frequently used in remote sensing for land cover classification. For instance, it can classify different types of vegetation, urban areas, and water bodies based on their spectral signatures. By computing the mean spectral values for each land cover type and comparing each pixel to these mean vectors, the algorithm effectively assigns each pixel to the appropriate class. So this algorithm is a valuable tool for image classification due to its simplicity and efficiency, though it requires careful consideration of its assumptions and limitations in scenarios such as complex class distributions or significant variations within classes.

14.4.2 Parallelepiped Classifier

Parallelepiped classifier uses the class limits stored in each class signature to determine if a given pixel falls within the class or not. The class limits specify the dimensions (in standard deviation units) of each side of a parallelepiped surrounding mean of the class in feature space. If pixel falls inside the parallelepiped, it is assigned to the class. However, if pixel falls within more than one class, it is put in the overlap class. If pixel does not fall inside any class, it is assigned to the null class i.e., left unclassified. Such pixels are taken up for classification at a later stage by examining the classes assigned to their immediate neighbours.

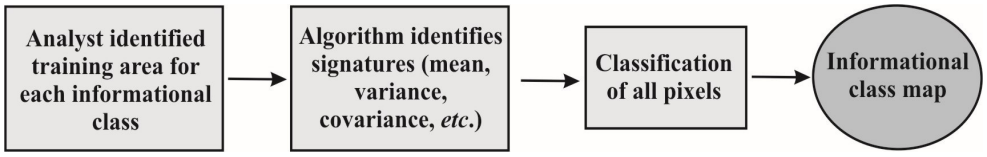

Fig. 14.9: Steps involved in supervised classification.

In parallelepiped classifiers, an n-dimensional imaginary box is constructed around pixels within each category of interest (Fig. 14.10). The n-dimensional space defined by the parallelepiped delimits different categories.

Classification using this classifier is carried out in the following major steps:

Step 1: Define the range of values in each training area and use these ranges to construct an n-dimensional box (a parallelepiped) around each class.

Step 2: Use multi-dimensional ranges to create different surface categories.

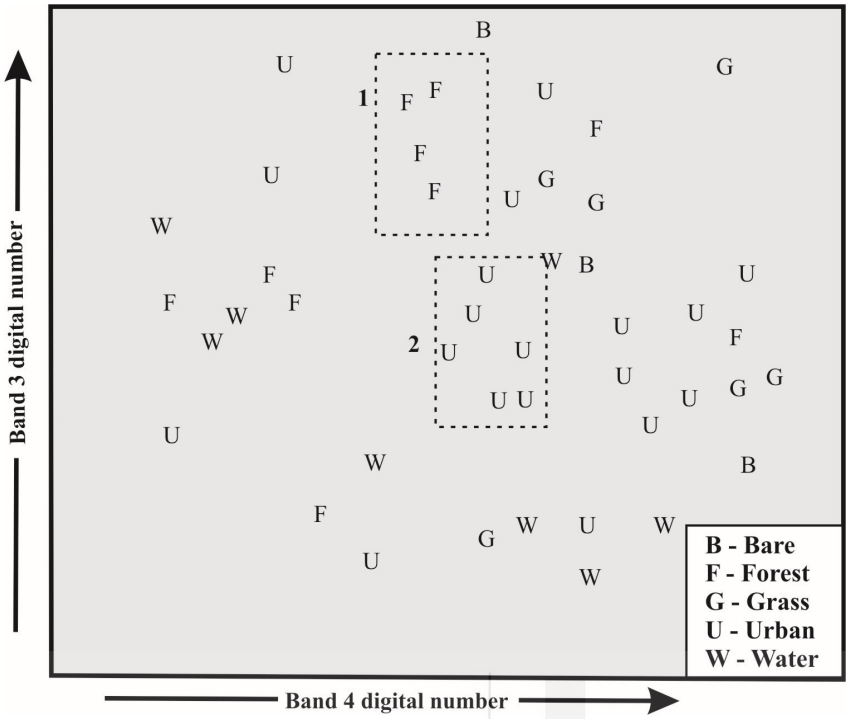

Fig. 14.10: Using the parallelepiped approach, pixel 1 is classified as forest and pixel 2 is classified as urban.

Notice that there can be overlap between the categories when simple method is used. One solution to this problem is to use a stepped decision region boundary.

Let us now see the advantages and limitations of this classifier.

Advantages

- it is a simple and computationally inexpensive method, and
- it does not assume a class statistical distribution and includes class variance.

Limitations

- it is least accurate method
- it does not adapt well to elongated (high-covariance) clusters
- it often produces overlapping classes, requiring a second classification step
- it also becomes more cumbersome with increasing number of channels, and
- pixels falling outside the defined parallelepiped remain unclassified.

14.4.3 Nearest Neighbour Classifier

The Nearest Neighbour (NN) algorithm is a commonly used supervised classification technique. It is based on the principle of classifying a sample based on the majority class of its nearest neighbour(s) in the feature space. In the context of image classification, it assigns labels to pixels based on the similarity between their feature vectors and those of labelled training samples. The classifier works by comparing the feature vector of an unknown data point (or pixel) to those of labelled training samples. As we have discussed earlier, the *distance* between the data points is typically measured using a distance

metric, such as Euclidean distance, and the data point is assigned to the class of the nearest neighbour(s).

There are following variants of the Nearest Neighbour classifiers:

- **1-Nearest Neighbour (1-NN):** This is the simplest version in which, a test point is assigned the class of its single closest training sample. It does not assume that the data is normally distributed.
- **k-Nearest Neighbours (k-NN):** In this version, instead of just one neighbour, the class label is determined by the majority vote from the k-nearest neighbours. This helps in making the classifier more robust to noise. It searches away from the pixel to be classified in all directions until it encounters *k* user specified training pixels and then assigns the pixel to the class with the majority of the pixels encountered (Jensen, 2018).
- **k-Nearest Neighbous distance-weighted (Weighted k-NN):** It uses the same *k*-nearest neighbours but it in this variation, closer neighbours are given higher weight when determining the class label. This means the nearest neighbours have a stronger influence on the classification than those farther away.

Classification using this classifier is carried out in the following major steps:

Step 1: Training Phase:

- **Collect training samples**: A labelled dataset is needed, where each data point (pixel or image patch) has a corresponding feature vector and class label.
- **Store the training samples**: The feature vectors and their corresponding class labels are stored for reference during the classification phase. No model training is actually done here; all samples are used as potential references for later classification.

Step 2: Classification Phase:

- **Select a test point**: The classifier receives an unknown data point (e.g., an image pixel) that needs to be classified.
- **Compute distances**: The distance between the test point's feature vector and the feature vectors of all training samples is calculated. Commonly used distance metrics are the following:
  - **Euclidean distance**: Most commonly used for continuous features, it measures the straight-line distance between two points in Euclidean space. It is calculated as: $d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$
  - **Manhattan distance**: Useful when movement is restricted to axes. It is calculated as $d(x, y) = \sum_{i=1}^{n} |x_i - y_i|$
  - **Cosine similarity**: Measures the angle between two vectors and is commonly used in text classification but can also be applied to image classification in some cases.
  - **Mahalanobis distance**: Takes into account correlations between variables, useful when the data points have correlated features.

- **Find the nearest neighbour(s)**: Identify the training sample(s) that are closest to the test point according to the distance metric chosen.
- **Assign class label**: The test point is assigned to the class that is most frequent among the nearest neighbours. In the **k-Nearest Neighbour (k-NN)** variant, the class label is determined by a majority vote from the *k* nearest neighbours.

Step 3: Output

The image pixel (or patch) is classified based on the most common class among its nearest neighbours.

Let us now see the advantages and limitations of this classifier.

Advantage:

- **Simplicity**: The algorithm is intuitive and easy to understand and implement.
- **No assumption**: It does not assume any underlying distribution for the data.
- **Non-linear boundaries**: These classifiers can handle complex decision boundaries between classes because the decision regions are based on proximity to neighbours, not predefined linear boundaries.
- **Locality-based classification**: It works well in cases where the class of a point is influenced by its immediate neighbours, making it suitable for spatially correlated data, like images.

Limitation:

- **Computational complexity**: It can be slow for large datasets, because the algorithm requires comparing the test point to every training sample.
- **Memory usage**: Since the algorithm stores all training samples for later classification, it can consume a lot of memory.
- **Sensitive to noise**: Nearest Neighbour classifiers are sensitive to noisy data or outliers. A noisy training sample can influence the classification of many other pixels.
- **Curse of dimensionality**: As the number of features (dimensions) increases, the distance between points becomes less meaningful. This issue, known as the "curse of dimensionality," can degrade the performance of NN classifiers, especially in high-dimensional feature spaces.
- **Imbalanced data**: In cases where some classes are significantly overrepresented, the classifier may be biased toward the dominant class unless specific measures are applied.

14.4.4 Maximum Likelihood Classifier

Maximum likelihood (MXL) classifier is one of the most widely used classifiers in the remote sensing. In this method, a pixel is assigned to the class for which it has maximum likelihood of membership. This classification algorithm uses training data to estimate mean vectors and band-wise variances of the classes, and pair-wise band to band covariances which are then used to estimate probabilities of pixels to belong to different classes. Maximum likelihood classification considers not only mean or average values in assigning

classification but also the variability of brightness values in each class around the mean. It is the most powerful of the classification algorithms as long as accurate training data is provided and certain assumptions regarding the distributions of classes are valid.

An advantage of this algorithm is that it provides an estimate of overlap areas based on statistics. This method is different from parallelepiped in that it uses only maximum and minimum pixel values. The distribution of data in each training set is described by a mean vector and a covariance matrix. Pixels are assigned a posteriori probability of belonging to a given class and placed in the most ‘‘likely’’ class. This algorithm takes into account the shape of the training set distribution.

For mathematical tractability, Maximum likelihood classifiers assume the conditional probability density of the pixel data vectors given the class (from the training data set) to be Gaussian (normal) distribution. With the help of an estimate of the proportion of each class within the study area, and the class conditional density function of the data vectors in the class, the probability of a class given the pixel data vector is computed and then as mentioned above, the pixel data vector is assigned to the most likely class whose conditional probability given the data vector is the highest.

The basis for the Gaussian assumption used above is that plotting the number of pixels with any given DN value yields a histogram or distribution of DN values within a particular band. Studies have shown that for most surfaces DN values from visible or near-infrared (NIR) region of the electromagnetic (EM) spectrum have a normal probability distribution. It means we can define curves based on the mean and standard deviation of the sample that describe the normal probability distribution by selecting category that has the highest statistical probability for each pixel. These concentric circles, called equi-probability contours, are derived from an assumed normal distribution around each training site. Equi-probability contours define the level of statistical confidence in the classification accuracy. Smaller the contour, higher is the statistical confidence.

Classification using this classifier is carried out in the following major steps:

Step 1: Training Phase

- **Collect training samples**: For each class (e.g., water, vegetation, urban), gather a set of labelled training samples that represent the spectral or feature characteristics of that class.
- **Compute class statistics**:
    - For each class, calculate the **mean vector** and **covariance matrix** from the training samples.
    - The mean vector (μ_C) represents the average feature values of the class.
    - The covariance matrix (Σ_C) represents the spread or variability of the features for that class.
    - These statistical values characterise the shape of the distribution of the class in the feature space.

      The **mean vector** for class C is computed as:

$$\mu_C = \frac{1}{N_C}\sum_{i=1}^{N_C} x_i$$

where,
x_i is a training sample for class C, and
N_C is the number of training samples for class C.

The **covariance matrix** for class C is calculated as:

$$\Sigma_C = \frac{1}{N_C - 1}\sum_{i=1}^{N_C}(x_i - \mu_C)(x_i - \mu_C)^T$$

where, $(x_i - \mu_C)$ is the deviation of the sample from the mean.

Step 2: Classification Phase

- **Compute the likelihood**: For each pixel in the image (or test sample), compute the likelihood that the pixel belongs to each class. This is done by evaluating the multivariate normal (Gaussian) distribution for each class using the class mean vector and covariance matrix. The likelihood function for class *C* is given by the multivariate Gaussian probability density function (PDF)**:**

$$P(x|C) = \frac{1}{(2\pi)^{d/2}|\Sigma_C|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_C)^T \Sigma_C^{-1}(x - \mu_C)\right)$$

where:
x is the feature vector of the pixel,
μ_C is the mean vector of class C.
Σ_C is the covariance matrix of class C,
d is the number of features (dimensions of the feature space), and
$|\Sigma_C|$ is the determinant of the covariance matrix.

- **Compute the posterior probability**: For each class, the posterior probability that a pixel belongs to class C is proportional to the likelihood of the pixel given the class, as well as the prior probability of the class. The posterior probability can be calculated using Bayes' theorem:

$$P(C|x) = \frac{P(x|C)P(C)}{P(x)}$$

where,
$P(C)$ is the prior probability of class C, and
$P(x)$ is the total probability of observing the pixel x (which acts as a normalisation constant and is the sum of the likelihoods over all classes).

Since the denominator $P(x)$ is the same for all classes, the classifier chooses the class with the highest posterior probability, which simplifies to maximising the likelihood $P(x|C)$.

- **Assign class labels**: After calculating the posterior probability for each class, the pixel is assigned to the class with the highest posterior probability.

Step 3: Output:

- The algorithm produces a classified image, where each pixel is assigned to a class based on the highest likelihood or posterior probability.

Let us now see the advantages and limitations of this classifier.

Advantage

- it is one of the most accurate methods
- it overcomes unclassified pixel problem (subject to threshold values)
- it provides a consistent way to separate pixels in overlap zones between classes and
- assignment of pixels to classes can be weighted by prior knowledge of the likelihood that a class is correct.

Limitation

- cluster distributions are assumed to be Gaussian in each class and band. Algorithm requires enough pixels in each training area to describe a normal population and assumes class covariance matrices are similar
- classes not assigned to training sets tend to be misclassified – a particular problem for mixtures
- it is reliant on the accuracy of training data. Changes in training set of any one class can affect the decision boundaries with other classes
- it is relatively computationally expensive, and
- it is also not practical with imaging spectrometer data.

14.4.5 Spectral Angle Mapper Classifier

The **Spectral Angle Mapper (SAM)** algorithm is a spectral-based supervised classification method, widely used in satellite and airborne remote sensing image analysis for land cover classification, vegetation analysis, mineral identification, and water quality monitoring. It is particularly effective in hyperspectral remote sensing where the spectral resolution is high. It is a classification technique that measures the spectral similarity between a given pixel's spectrum and the reference spectrum of a class. The basic idea behind SAM is to compare the angles between spectral vectors rather than the raw spectral values, which helps to mitigate the effects of illumination changes and sensor calibration issues. The SAM algorithm treats the spectral data as vectors in a multi-dimensional space (where each dimension corresponds to a spectral band). The spectral angle between the feature vector of a pixel and the reference spectrum of a class is computed. The lower the angle, the more similar the two spectra are, and thus, the pixel is more likely to belong to that class.

Classification using this classifier is carried out in the following major steps:

Step 1: Training Phase:

- **Collect training samples**: For each class, select a set of representative samples (training pixels) that define the class's spectral characteristics.
- **Create class mean spectra**: The reference spectrum for each class is typically calculated as the mean spectrum of all the training samples belonging to that class.

Step 2: Classification Phase:

- **Calculate spectral angle**: For each pixel (or test sample) in the image, calculate the spectral angle between the pixel's spectrum and each of the class reference spectra using the formula:

$$\theta = \cos^{-1}\left(\frac{\mathbf{X} \cdot \mathbf{Y}}{|\mathbf{X}||\mathbf{Y}|}\right)$$

where:
$\mathbf{X}$ is the spectrum of the pixel being classified,
$\mathbf{Y}$ is the reference spectrum of a class,
- denotes the dot product,
$|\mathbf{X}|$ and $|\mathbf{Y}|$ are the magnitudes (Euclidean norms) of the vectors $\mathbf{X}$ and $\mathbf{Y}$, respectively.

- **Assign class label**: The pixel is assigned to the class whose reference spectrum has the smallest spectral angle to the pixel's spectrum. This corresponds to the class with the most similar spectral signature.

Step 3: Output:

- The algorithm produces a classified image where each pixel is assigned to a class based on its spectral similarity to the reference spectra.

Let us now see the advantages and limitations of this classifier.

Advantage:

- Insensitive to signal intensity: Unlike other distance-based classifiers (e.g., Euclidean distance), SAM is insensitive to the overall intensity of the spectrum (e.g., illumination or reflectance variations). It only focuses on the spectral shape, making it robust to varying illumination conditions.
- Effective for hyperspectral data: SAM is well-suited for hyperspectral images with high-dimensional spectral data, as it compares spectral vectors directly, making it effective in distinguishing between materials that have similar reflectance patterns in specific spectral bands.
- Simplicity: The algorithm is conceptually simple and easy to implement, involving basic vector operations.
- **Robust to atmospheric effects**: SAM is less affected by atmospheric effects and radiometric differences than traditional classifiers because it compares the shape of the spectra rather than absolute values.
- **Handling of endmembers**: SAM is especially useful when dealing with mixed pixels, such as in vegetation or urban areas, as it can distinguish materials based on their spectral signatures without needing a physical model of the pixel mixture.

Limitation:

- **Sensitivity to noise**: SAM is sensitive to noise in the data. Small variations in the pixel spectrum or errors in the reference spectra can lead to misclassification.
- **No Handling of spatial context**: SAM classifies each pixel individually based on its spectrum, ignoring spatial information and context. This can lead to misclassifications, especially in homogeneous or patchy areas.

- **Non-optimal for highly similar classes**: SAM might struggle when class separability is low, especially when classes have very similar spectral signatures. In such cases, even small spectral differences can lead to errors in classification.
- **Training data requirement**: SAM is a supervised classification method, meaning it requires a sufficient amount of representative training data for each class. If the training data is inadequate or unrepresentative, the classifier's performance will degrade.
- **Dimensionality concerns**: In the case of hyperspectral images, the dimensionality of the data can be very high. SAM operates by calculating angles in high-dimensional spaces, which could be computationally expensive and may suffer from the curse of dimensionality unless dimensionality reduction techniques (e.g., PCA) are used.

14.4.6 Artificial Neural Network Classifier

Neural network classifiers do not require any assumption about the statistical distribution of data and use machine learning techniques to classify image pixels. **Artificial Neural Networks (ANNs)** are a class of machine learning algorithms inspired by the structure and function of the human brain. In supervised image classification, ANNs are used to learn complex patterns in the data by simulating a network of interconnected nodes (neurons). Each neuron performs simple computations, and the network as a whole can recognise intricate patterns and relationships in the input data. ANNs are particularly useful in tasks like image classification, object recognition, and semantic segmentation because they can model non-linear relationships between input features.

ANNs consist of layers of interconnected neurons that process input data. The most common ANN used for image classification is the **Multilayer Perceptron (MLP)**, which is a feedforward neural network.

Classification using this classifier is carried out in the following major steps:

Step 1: Training Phase:

- **Data Preparation**:
  - A labelled dataset is needed for training the network, where each image (or image patch) is associated with a known class label (e.g., vegetation, water, buildings).
  - Preprocess the data by normalising or scaling the input features, such as pixel intensities, to a standard range (e.g., [0,1] or [-1, 1]).
- **Network Initialisation**:
  - Initialise the ANN architecture, which typically consists of three layers: an **input layer**, one or more **hidden layers**, and an **output layer**.
  - The **input layer** represents the features (e.g., pixel values or extracted image features), and the **output layer** contains one neuron for each class, representing the class probabilities.
- **Forward Propagation**:
  - Pass the input features through the network. For each neuron in each layer:

- Compute the weighted sum of inputs and add a bias term.
- Apply an activation function (such as sigmoid, ReLU, or tanh) to introduce non-linearity.
- The output of the last layer represents the network's prediction for each class (often using a softmax function for multi-class problems).

- **Loss Function Calculation**:
  - Calculate the loss (or error) by comparing the predicted outputs with the true class labels using a loss function, such as cross-entropy loss for classification tasks.

$$L = -\sum_i y_i \log(\hat{y}_i)$$

where, y_i is the true label and $\hat{y}_i$ is the predicted probability for class *i*.

- **Backpropagation**:
  - Compute the gradients of the loss with respect to each weight in the network using the chain rule. This process is called backpropagation and allows the network to adjust the weights.
  - Update the weights using an optimisation algorithm, such as Gradient Descent or its variants (e.g., Adam, RMSprop), to minimise the loss function.
- **Iteration**:
  - Repeat the process for multiple epochs (iterations over the entire dataset), adjusting the weights until the model converges to a state where the error is minimised.

Step 2: Classification Phase:

- **Prediction**:
  - For a new, unseen image or pixel, pass the input features through the trained neural network (forward propagation).
  - The network will output class probabilities, and the class with the highest probability is selected as the predicted class label.
- **Output**:
  - The result is a classified image where each pixel is assigned to one of the predefined classes based on the trained neural network model.

There are following variants of ANNs used in image classification:

1. **Multilayer Perceptron (MLP)**:
   - The most basic form of ANN with fully connected layers.
   - It is typically used for structured data classification but can also be applied to image classification if the image is flattened into a vector.
2. **Convolutional Neural Networks (CNNs)**:
   - A specialised variant of ANNs designed specifically for image classification.
   - CNNs use convolutional layers to automatically learn spatial hierarchies of features (such as edges, textures, and object parts), making them very effective for image data.

- They typically involve pooling layers for downsampling, activation layers for non-linearity, and fully connected layers for final classification.

3. **Recurrent Neural Networks (RNNs)**:
   - While not commonly used in standard image classification tasks, RNNs can be applied to sequential image data, such as video frames, where temporal dependencies are important. They process data in a time-dependent manner and are particularly useful for tasks like image captioning or video analysis.
4. **Deep Neural Networks (DNNs)**:
   - A type of neural network with many hidden layers. They have greater representational power compared to traditional shallow neural networks and can model highly complex patterns in data.
   - DNNs are often used as the foundation for more advanced architectures like **CNNs** or **Generative Adversarial Networks (GANs)**.

Let us now see the advantages and limitations of this classifier.

Advantage:

- Non-linearity: ANNs can model highly complex, non-linear relationships between the input and output, which makes them highly effective in capturing complex patterns in image data.
- Feature learning: In the case of CNNs, the network automatically learns relevant features from raw pixel data without the need for manual feature extraction.
- Adaptability: ANNs are highly adaptable and can be used for a wide range of classification tasks, from simple pixel-based classification to complex object recognition and segmentation.
- Ability to handle high-dimensional data: ANNs, particularly CNNs, are well-suited for handling high-dimensional data, such as hyperspectral images or images with a large number of bands.
- Robustness: ANNs are generally robust to noise and small variations in the input data, especially when trained with a large and diverse dataset.

Limitation:

- Data requirement: ANNs typically require large amounts of labeled data to train effectively, especially deep networks like CNNs. For small datasets, they can easily overfit.
- Computational complexity: Training deep neural networks can be computationally expensive and time-consuming, requiring powerful hardware (e.g., GPUs) and careful optimisation techniques.
- Interpretability**:** ANNs, especially deep networks, are often considered "black boxes" because it can be difficult to understand why they make certain decisions, limiting their use in applications where interpretability is crucial.
- Risk of overfitting: Without proper regularisation (e.g., dropout, early stopping) or a sufficiently large training dataset, ANNs can easily overfit, learning the noise in the data rather than the true underlying patterns.
- Hyperparameter tuning: The performance of ANNs depends on the correct choice of hyperparameters, such as the learning rate, number of layers, number of neurons per layer, and activation functions. Tuning these parameters is often a trial-and-error process.

ANNs have revolutionised image classification tasks, especially with the advent of Convolutional Neural Networks (CNNs). They excel at handling complex, high-dimensional data and learning hierarchical features, making them ideal for tasks such as image analysis**,** object recognition**,** and also medical imaging. CNNs are particularly effective for this task. In remote sensing, they are used for classifying land use land cover types, detecting deforestation or any objects, identifying water bodies, and classifying vegetation types in satellite and aerial images. They are also used in identifying objects in images, such as cars, pedestrians, and animals and even for facial recognition and self driving cars. In medical imaging they are used for classifying tissues, lesions, or abnormalities in MRI, CT scans, or X-ray images and for automated diagnosis and disease detection. While they offer significant advantages in terms of adaptability and performance, they come with challenges related to data requirements, computational resources, and interpretability. Despite the limitation, ANNs remain a cornerstone of modern image classification techniques due to their ability to model intricate patterns in data effectively.

14.4.7 Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm widely used for remote sensing image classification. It is a non-parametric statistical learning technique that is particularly effective for high-dimensional data with limited training samples. SVM aims to find an optimal hyperplane that separates different classes in a feature space with the maximum margin. Due to its ability to handle complex and non-linear class distributions, SVM has become a popular choice for land cover classification in remote sensing.

Classification using this classifier is carried out in the following major steps:

Step 1: Data Preparation

- Collect training samples representing different land cover classes.
- Extract spectral, spatial, or textural features from the remote sensing image.
- Normalise or standardise features to ensure better model performance.

Step 2: Kernel Selection

- Choose an appropriate kernel function to transform the data into a higher-dimensional space where classes are more separable. Common kernel functions include:
  - Linear Kernel (for linearly separable data).
  - Polynomial Kernel (for more complex patterns).
  - Radial Basis Function (RBF) Kernel (for non-linearly separable data).
  - Sigmoid Kernel (similar to neural networks).

Step 3: Training the SVM Model

- The SVM algorithm identifies support vectors, which are the most informative training samples defining the class boundaries.
- It constructs an optimal hyperplane that maximises the margin between classes.

Step 4: Hyperparameter Optimisation

- Tune parameters such as C (regularisation parameter) and γ (gamma for RBF kernel) to balance complexity and generalisation.
- Cross-validation is used to determine the best parameter values.

Step 5: Classification and Label Assignment

- Each pixel in the remote sensing image is classified based on its position relative to the hyperplane.
- The model assigns class labels based on the learned decision boundary.

Step 6: Accuracy Assessment

- Perform an accuracy assessment using techniques like confusion matrix, overall accuracy, producer's accuracy, and user's accuracy.
- Compare SVM results with ground truth or reference datasets.

Let us now see the advantages and limitations of this classifier.

Advantage

- **Effective in high-dimensional spaces**: Works well with hyperspectral and multispectral imagery where the number of features is large.
- **Good performance with small training samples**: Unlike neural networks, SVM does not require a large training dataset for effective classification.
- **Handles non-linear data separability**: By using kernels, SVM can model complex class boundaries in remote sensing images.
- **Robust to overfitting**: Especially when using an appropriate kernel and regularisation parameters.
- **Versatile and generalisable**: Can be applied to different types of remote sensing data, including optical, radar, and LiDAR images.

Limitation

- **Computationally intensive**: Training can be slow for large datasets, especially with non-linear kernels.
- **Difficult parameter selection**: Requires careful tuning of kernel parameters (e.g., C and γ) to achieve optimal performance.
- **Sensitive to noisy data and overlapping classes**: Misclassification can occur in highly mixed or noisy environments.
- **Binary classifier by default**: Requires some extensions for multi-class classification.

SVM is considered as a powerful and widely used classifier in remote sensing due to its ability to handle high-dimensional data and complex decision boundaries. Despite its computational challenges, SVM remains a preferred choice for many classification tasks, particularly when training data is limited. With proper kernel selection and parameter tuning, SVM can achieve high accuracy in remote sensing image classification.

14.4.8 Random Forest Classifier

Random Forest (RF) is a widely used supervised machine learning algorithm in remote sensing image classification. It is an ensemble learning method that constructs multiple decision trees and combines their outputs to improve classification accuracy and reduce overfitting. RF is known for its robustness, efficiency, and ability to handle high-dimensional remote sensing data, including multispectral, hyperspectral, and LiDAR datasets.

Classification using this classifier is carried out in the following major steps:

Step 1: Data Preparation

- Collect training samples representing different land cover classes.
- Extract spectral, spatial, or textural features from the remote sensing image.
- Split the dataset into training and testing sets.

Step 2: Building Multiple Decision Trees

- **Bootstrap Sampling**: The RF algorithm randomly selects subsets of training data with replacement (bootstrap sampling).
- **Feature Selection for Each Tree**: A random subset of features is chosen for each decision tree to reduce correlation among trees.
- **Tree Construction**: Each decision tree is trained independently using the selected subset of data and features.

Step 3: Classification and Majority Voting

- Each decision tree classifies the pixels in the image independently.
- The final class label is assigned based on the majority vote of all decision trees.

Step 4: Model Validation and Accuracy Assessment

- Evaluate the classification results using a confusion matrix to calculate accuracy metrics such as overall accuracy, producer's accuracy, and user's accuracy.
- Perform cross-validation to assess the generalisation capability of the model.

Let us now see the advantages and limitations of this classifier.

Advantage

- **High accuracy and robustness**: RF reduces overfitting and provides stable results, even with complex land cover classes.
- **Handles high-dimensional data well**: Suitable for hyperspectral, multispectral, and LiDAR data where many spectral and spatial features exist.
- **Works well with small training samples**: Unlike deep learning models, RF performs well even with limited training data.
- **Insensitive to noisy data and missing values**: RF is less sensitive to noise and can handle missing data effectively.
- **Fast and parallel processing**: Training and classification processes are relatively fast due to parallel computation of decision trees.

Limitations

- Computational complexity**:** Requires more memory and processing power when handling very large datasets.
- Not optimal for highly correlated features: If many features are correlated, some trees may become redundant, reducing the model's efficiency.
- Difficult interpretation**:** Unlike single decision trees, RF does not provide an easily interpretable classification rule.
- Bias toward majority classes: If the training dataset is imbalanced, RF may favour majority classes, leading to misclassification of minority classes.

Random Forest is a powerful and widely used classification algorithm in remote sensing due to its high accuracy, robustness, and ability to handle complex da-

tasets. Its ability to reduce overfitting and work with high-dimensional features makes it a preferred choice for land cover classification, vegetation mapping, and environmental monitoring. Despite some computational challenges, RF remains one of the most reliable classifiers for remote sensing image analysis.

We have become familiar with many algorithms of supervised classification. Let us compare and summarise what we have read in this section. Table 14.2 summarises the various classifiers we have read:

Table 14.2: Comparison of the commonly used supervised classification algorithms:

Aspect	Minimum Distance to Means	Parallelepiped Classifier	Nearest Neighbour Classifier	Maximum Likelihood Classifier	Spectral Angle Mapper Classifier	Artificial Neural Network Classifier	Support Vector Machine	Random Forest Classifier
Description	Classifies based on the closest class mean in feature space	Uses a box decision rule based on class-specific ranges	Classifies based on the majority class of nearest neighbours	Uses probability density functions to assign classes	Measures spectral similarity using angle between vectors	Uses interconnected neurons to model complex patterns	Maximises the margin between classes using hyper-planes	Uses ensemble of decision trees for classification
Accuracy	Moderate	Moderate	High	High	High	Very High	Very High	Very High
Processing Time	Fast	Fast	Slow	Moderate	Moderate	Slow	Slow	Slow
Complexity	Low	Low	Moderate	High	High	High	High	High
Resource Usage	Low	Low	Moderate	High	Moderate	High	High	High
Suitability	Simple data	Simple data	Complex data	Complex data	High-dimensional data	Complex patterns	Complex patterns	Complex data
Flexibility	Low	Low	High	Moderate	High	Very High	High	High
Handling of Mixed Pixels	Poor	Poor	Good	Good	Good	Very Good	Very Good	Very Good
Interpretability	Easy	Easy	Moderate	Moderate	Moderate	Difficult	Difficult	Moderate
Advantage	Simple and efficient	Easy to implement	High accuracy, adaptable	High accuracy, considers class probabilities	Robust to variations in spectral signatures	Models complex relationships, high accuracy	Handles non-linear relationships, high accuracy	Robust, handles large datasets
Limitation	Sensitive to outliers, assumes linear separability	Sensitive to range definitions, limited flexibility	Slow for large datasets, sensitive to noise	Assumes normal distribution, computationally intensive	Sensitive to spectral variability	Requires large training data, computationally intensive	Requires tuning, computationally intensive	Requires tuning, computationally intensive

14.5 OVERVIEW OF OTHER CLASSIFICATION METHODS

As mentioned in the above sections, apart from the supervised and unsupervised classification there exists a third kind of classification method known as **hybrid classification**. This method uses both the afore-mentioned methods (supervised and unsupervised classifications) and is primarily applied to improve accuracy and efficiency of classification results. The most common example of hybrid classification is the use of unsupervised classification to delineate classes prior to supervised classification in order to aid the analyst in

identifying numerous spectral classes. Guided clustering is another such method which is useful in analysis involving complex variability in spectral response for each land cover. In this method, analysis delineates numerous supervised training sets for each land cover. These training sets do not have to be homogeneous as opposed to the regular supervised classification. Data from all these training sets are used for supervised classification. The analyst uses his discretion while selecting final spectral classes, so all of the redundant classes are merged or discarded as per need.

There are number of other classification methods such as contextual, decision tree, multisource, etc.

Remote sensing images often contain complex landscapes where similar land cover types may have varying spectral signatures due to environmental conditions, sensor characteristics, or mixed pixels. Traditional classifiers, such as Maximum Likelihood Classification (MLC) and Support Vector Machines (SVM), may struggle with classifying such images accurately. Contextual classification addresses these challenges by integrating spatial dependencies and improving classification accuracy, particularly in heterogeneous landscapes. **Contextual classifiers** incorporate spatial or temporal information along with the spectral signatures while deciding the information classes. Contextual classifiers have been developed to address intraclass spectral variations by incorporating spatial information among neighbouring pixels, thereby improving classification accuracy. These classifiers utilise techniques such as smoothing, Markov random field, spatial statistics, fuzzy logic, segmentation, and neural networks. In short, Markov Random Fields (MRF) is a probabilistic approach that models the spatial dependencies between neighbouring pixels. It assumes that the class of a pixel depends on its neighbours, improving classification accuracy through Bayesian inference. Pre-smoothing approaches integrate contextual information as additional bands before spectral classification, while post-smoothing is applied to already classified images. Among these methods, Markov random field-based classifiers, particularly iterated conditional modes, are widely used and have proven effective in enhancing classification results. Advantages of contextual classification include: higher accuracy by reducing misclassification errors in heterogeneous regions; better handling of mixed pixels by considering spatial relationships; improved representation of land cover by incorporating shape, texture, and neighbourhood properties; and reduced salt-and-pepper noise compared to pixel-based classification. However, it is computationally intensive, requiring significant processing power and memory and is dependent on high-resolution data, making it less effective for coarse-resolution imagery.

Decision tree classifiers are knowledge-based classifiers which classify in steps, where the classifier is able to distinguish between two or more classes at each step. In this method, various types of classifiers as deemed appropriate can be combined. A Decision Tree (DT) classifier is a supervised machine learning algorithm used for remote sensing image classification. It operates by recursively splitting data into subsets based on feature values, forming a tree-like structure where each node represents a decision rule and each leaf represents a class label. Major steps in the Decision Tree classification involve data collection and preprocessing; tree construction as the algorithm selects the

best features to split the dataset at each node using some criteria; recursive partitioning – which continues until stopping conditions (e.g., pure nodes or a maximum tree depth) are met; classification and prediction as new pixels are classified by traversing the tree from root to leaf nodes based on their feature values. Decision Tree (DT) classifier has several advantages such as easy to understand and visualise classification rules; effectively models complex land cover distributions; works well with large datasets; and there is no assumption about data distribution unlike parametric classifiers (e.g., Maximum Likelihood). However, it has some limitations such as it is prone to overfitting as deep trees may capture noise instead of patterns; small changes in data can alter the tree structure and could be biased towards dominant classes.

The incorporation of ancillary data, such as digital elevation models, soil maps, population density, and climate data, enhances remote sensing classification accuracy. **Knowledge-based classification** leverages spatial distribution patterns of land cover and ancillary data to refine classification results. For example, elevation and slope help differentiate vegetation types in mountainous regions, while population and road density aid in distinguishing urban land-use classes. Climate and soil data further improve large-scale land-cover mapping. A key step is developing classification rules, which can be created through expert knowledge, cognitive methods, or empirical data-driven techniques. GIS plays a crucial role in managing multisource data and supporting knowledge-based classification through spatial modelling.

Different classifiers, such as parametric (e.g., Maximum Likelihood) and non-parametric (e.g., Neural Networks, Decision Trees), have distinct strengths and limitations. MLC performs well when training samples are sufficient and data follows a normal distribution, whereas non-parametric classifiers handle irregular data distributions better. Research suggests that integrating multiple classifiers enhances classification accuracy compared to using a single classifier. Effective integration requires suitable combination rules, such as production rules, sum rules, stacked regression, majority voting, and threshold-based methods. **Hybrid classification** is an advanced approach in remote sensing that combines multiple classification techniques to improve accuracy and reliability in land use and land cover mapping. It integrates the strengths of different methods, such as supervised and unsupervised classification, pixel-based and object-based classification, or machine learning and rule-based approaches, to enhance classification performance, reduce misclassification errors, handle spectral variability, and effectively distinguish complex land cover types. This technique is particularly useful in complex landscapes where a single classification method may not be sufficient to distinguish between spectrally similar classes. Hybrid classification is a powerful approach that enhances the accuracy and reliability of remote sensing image classification by integrating multiple methods. While it requires advanced processing and expertise, its ability to combine spectral, spatial, and contextual information makes it highly effective for various remote sensing applications.

Multisource classification refers to the process of integrating and analysing data from multiple sources, such as optical, radar, LiDAR, hyperspectral, and ancillary data (e.g., DEMs, climate, and socioeconomic data), to improve classification accuracy. Unlike traditional single-source classification,

multisource approaches leverage the complementary nature of different datasets to enhance feature discrimination and reduce uncertainties. This technique is widely used in land use land cover classification, vegetation mapping, disaster monitoring, and urban planning, where different remote sensing data types provide varied and valuable insights.

Some classification methods deal with the classification of mixed pixels, which refer to pixels having classes corresponding to more than one land cover. These methods, referred to as spectral mixture analysis, are based on physical models providing information on discrete spectral signatures rather than statistical methods.

Linear mixture methods consider spectral classes from one pixel to be linear mixture of all the land cover classes. As you have read in the previous unit, fuzzy classification methods account for the transition between various land cover classes, known as fuzzy regions in between two classes. Fuzzy classification does not have definite boundaries and one pixel may belong to more than one class.

14.6 ROLE OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN IMAGE CLASSIFICATION

You have often come across the terms like artificial intelligence, machine learning and deep learning, which are at times used interchangeably. However, there is some difference. Artificial intelligence (AI) is used to classify machines that mimic human intelligence and human cognitive functions like problem-solving and learning. Artificial intelligence is an umbrella term under which the machine learning (that allows for optimisation) and deep learning come. While machine learning (ML) is a subset of artificial intelligence, deep learning is a subfield of machine learning. Neural networks form the backbone of deep learning algorithms. Deep learning has much higher number of node layers, or depth, of neural networks than a single neural network.

Let us now try to understand the role and importance of AI and ML in image classification. It may be useful for the following tasks:

- **Enhanced Accuracy and Performance**: these are good in recognising complex patterns and features in images and can automatically learn details and subtle differences between classes, leading to higher classification accuracy.
- **Adaptability to Various Data Types**: the machine learning models can be trained to handle different types of image data, including multispectral, hyperspectral, and radar images ensuring that classification tasks are accurate across various types of images acquired from different types of sensors.
- **Handling High-Dimensional Data**: these techniques can manage and process vast amount of spectral bands efficiently without requiring manual feature extraction.

- **Automated Feature Extraction**: while traditional methods often require manual feature extraction, which can be time-consuming and error-prone, the AI and ML can be used to automate feature extraction, making the classification process faster and more consistent.
- **Scalability**: AI and ML algorithms can be scaled to handle large volumes of datasets quickly, facilitating real-time and large-scale image classification.
- **Robustness to Variability**: ML models are robust to variations in data such as changes in illumination, sensor differences, and also varying spatial resolutions and can generalise well across diverse datasets, reducing the impact of such variations on classification accuracy.
- **Probabilistic and Soft Classification**: AI techniques like ensemble methods and Bayesian networks provide probabilistic outputs, offering insights into the uncertainty associated with classification results which is particularly useful for decision-making for disaster management and environmental monitoring.
- **Object Detection**: AI methods, particularly object detection algorithms like YOLO (You Only Look Once) and Faster R-CNN, can identify and localise objects within an image, which is an important capability for applications such as urban planning, infrastructure monitoring, and wildlife tracking and even for military applications.
- **Change Detection**: these can detect and analyse changes over time by comparing classified images from different periods which can be useful for monitoring deforestation, urban sprawl, and other dynamic processes.
- **Continuous Improvement and Learning**: these models can be updated and retrained as new data becomes available, allowing them to adapt to changes in data distributions and improve over time. This continuous learning process ensures that the models remain effective as conditions evolve.

AI and ML have revolutionised image classification by providing advanced techniques for accurate, scalable, and efficient analysis of complex image data. They automate feature extraction, handle high-dimensional data, and adapt to varying conditions, making them indispensable in modern remote sensing and computer vision applications. As technology continues to advance, the capabilities and applications of AI and ML in image classification are expected to expand further, offering even more powerful tools for interpreting and understanding various kinds of images.

14.8 SELECTION OF APPROPRIATE CLASSIFICATION METHOD

Classification process involves translating pixel values in a remote sensing image into meaningful categories. In case of land cover classification, these categories comprise different types of land cover defined by the classification scheme that is being implemented. There are number of classification methods that can be used to group image pixels into meaningful categories.

Unfortunately, there is not a single best approach to image classification. The choice made depends a lot on the algorithms that are available with the image processing software used and also familiarity and experience with different methods. The choice of classification method is dependent upon many factors, accuracy being one of the most important criteria. Some of the ways to evaluate efficacy and accuracy of classification methods are discussed in next unit. It has generally been found that in areas of complex terrain, the unsupervised approach is preferable over the supervised one. In such conditions, if the supervised approach is used, the user will have difficulty in selecting training sites because of the variability of spectral response within each class.

Consequently, a prior ground data collection can be very time consuming. Also, the supervised approach is subjective in the sense that the analyst tries to classify information categories, which are often composed of several spectral classes whereas spectrally distinguishable classes will be revealed by the unsupervised approach, and hence ground data collection requirements may be reduced. Additionally, unsupervised approach has the potential advantage of revealing distinguishable classes unknown from previous work. However, when definition of representative training areas is possible and statistical information classes show a close correspondence, the results of supervised classification will be superior to unsupervised classification.

Fig. 14.11 shows a Landsat scene of Washington DC, USA which has been classified by supervised as well as unsupervised methods. We can observe that there are many similarities between outputs of supervised and unsupervised methods. However, in this example the outcome of the supervised classification method has more generalised classes than that of the unsupervised method.

This is because of the fact that while we use spectral information to create classes in the unsupervised classes, performance of the supervised classes is largely dependent on the training samples. The training samples used for the above classification (Fig. 14.11) is not sufficient to cover the entire spectrum of a particular class and, therefore, we get a generalised image. The results of the supervised classification can be further improved by collecting more training samples which would further help to reduce the differences between classes which may be due to mixtures within each pixel e.g. grass and forest.

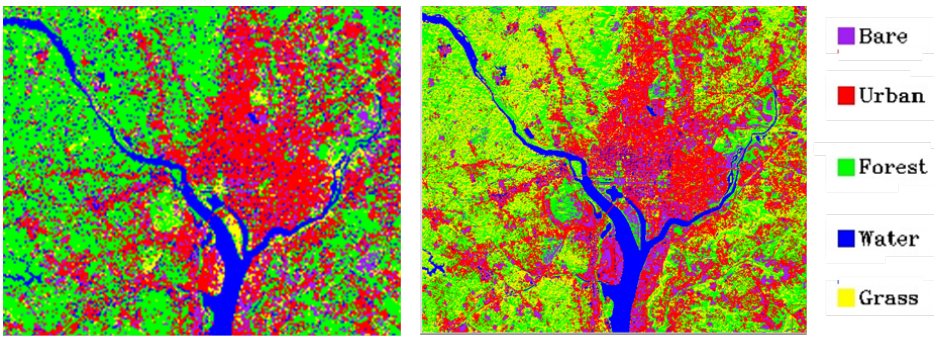

Fig. 14.11: Processing of images: a) Supervised classification; and b) Unsupervised classification with five classes.

It is not easy to answer the question which classification method is suitable for a study because different classification results can be obtained from different methods and each method has its own merits and demerits. However, for a

general guideline it can be said that factors such as spatial resolution of the remote sensing data, source of data, classification scheme and availability of classification software must be taken into account while selecting a classification method for use.

Let us spend 5 minutes to check your progress.

SAQ II

a) What is minimum distance to means algorithm?

b) What is nearest neighbour classifier?

c) What is contextual classification?

d) What is multisource classification?

14.8 RECENT TRENDS

Recent trends and developments in remote sensing image classification have been driven by advancements in computational capabilities, machine learning algorithms, data availability, and also sensor technologies. These innovations have significantly enhanced the accuracy, efficiency, and applicability of classification methods across various fields. Some of these are listed here:

- Deep Learning and Convolutional Neural Networks (CNNs) for automatically learning hierarchical features from data. Multi-Scale CNNs is an important networks which can capture information at different scales..
- Hybrid models combining deep learning with other machine learning techniques (e.g., Random Forest, SVM) are being developed to enhance classification accuracy and overcome individual method limitations.
- Integration of Object-Based Image Analysis (OBIA) and Geospatial AI for improved object recognition and classification.
- Multi-Source and Multi-Temporal Data Fusion to integrate various data types to enhance classification by leveraging complementary information, reducing uncertainty, and improving robustness.
- Use of cloud platforms such as Google Earth Engine (GEE) and AWS are increasingly being used as they provide powerful tools for processing and analysing large-scale remote sensing data.
- Automated and Self-Supervised Learning Approaches are being developed for large-scale classification, anomaly detection, and urban feature extraction.
- Real-Time and Near-Real-Time Classification is another emerging trend which could be useful for disaster response, traffic monitoring, and dynamic environmental assessments.

14.8 SUMMARY

Let us summarise what we have learnt in this unit. We have learnt that:

- There are two approaches to image classification, i.e. unsupervised and supervised. Unsupervised classification is useful for complex terrains and it can significantly reduce the cost of ground data collection than the supervised classification.
- Spectral signatures are unique for each material, which allows us to distinguish objects from one another and is the basis of classification in remote sensing.
- Methods of supervised image classification include Minimum distance to means, Nearest neighbour, Parallelepiped, maximum likelihood, SVM, SAM, ANN, RF, etc.
- Maximum likelihood method is the most powerful of the classification methods as long as accurate training data is provided and normal distribution of classes is justified. Sometimes, for better image classification, both supervised and unsupervised methods may be used which is known as a hybrid approach.
- Selection of an appropriate image classification method is a challenging task in image classification because there are so many classification methods available. However, the choice for a particular classification method depends on the availability of image processing software along with familiarity and working experience with other methods.
- There are several important developments in the field.

14.9 TERMINAL QUESTIONS

1. What is classification scheme?
2. What is training data?
3. What are spectral signatures?
4. What are the recent trends in supervised classification?
5. What is the role of AI and ML in image classification?
6. Give a comparison of various supervised classification algorithms.
7. Give a comparison of various statistical separability measures.

14.10 REFERENCES

- Abburu, S. & Golla, S. B. (2015). Satellite image classification methods and techniques: A review. International Journal of Computer Applications, 119(8).
- Bishop, C.M. (2006). Pattern recognition and machine learning by Christopher M. Bishop. Springer Science+ Business Media, LLC.
- Dhingra, S. & Kumar, D. (2019). A review of remotely sensed satellite image classification. International Journal of Electrical and Computer Engineering, 9(3), 1720.

- ERDAS Field Guide, http://www.gis.usu.edu/unix/imagine/FieldGuide.pdf (data retrieved in February, 2012).
- Jensen, J.R. (1986). Introductory Digital Image Processing: A Remote Sensing Perspective, Prentice-Hall, New Jersey.
- Lillesand, T. M., Kiefer, R. W. & Chipman, J. W. (2015). Remote Sensing and Image Interpretation. John Wiley & Sons.
- Lillesand, T.M. and Kiefer, R.W. (1994). Remote Sensing and Image Interpretation, John Wiley and Sons, Hoboken.
- Liu, J., & Mason, P. J. (2009). Essential Image Processing and GIS for Remote Sensing. John Wiley & Sons.
- Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. International journal of Remote sensing, 28(5), 823-870.
- Mather, P. M. & Tso, B. (2016). Classification Methods for Remotely Sensed Data. CRC Press.
- Richards, J. A. & Jia, X. (2006). Remote Sensing Digital Image Analysis: An Introduction. Springer.
- Salah, M. (2017). A survey of modern classification techniques in remote sensing for improved image classification. Journal of Geomatics, 11(1), 21.
- Swain, P.H. and Davis, S.M. (1978). Remote Sensing: The Quantitative Approach, McGraw-Hill International Book Co., New York.
- US Geological Survey Spectral Library, http://speclab.cr.usgs.gov/spectral-lib (data retrieved in September, 2011).
- Wilkie, D.S. and Finn, J.T. (1996). Remote Sensing Imagery for Natural Resources Monitoring – A Gide for First-Time Users, Columbia University Press, New York.

14.11 FURTHER/SUGGESTED READINGS

- Campbell, J.B. (2006). Introduction to Remote Sensing, Taylor and Francis, London.
- Lillesand, T. M., Kiefer, R. W. & Chipman, J. W. (2015). Remote Sensing and Image Interpretation. John Wiley & Sons.
- Lillesand, T.M. and Kiefer, R. (2007). Remote Sensing Image Interpretation, John Wiley, New York.
- Lillesand, T.M. and Kiefer, R.W. (1994). Remote Sensing and Image Interpretation, John Wiley and Sons, Hoboken.
- Liu, J., & Mason, P. J. (2009). Essential Image Processing and GIS for Remote Sensing. John Wiley & Sons.
- Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. International journal of Remote sensing, 28(5), 823-870.

- Mather, P. M. & Tso, B. (2016). Classification Methods for Remotely Sensed Data. CRC Press.
- Richards, J. A. & Jia, X. (2006). Remote Sensing Digital Image Analysis: An Introduction. Springer.
- Salah, M. (2017). A survey of modern classification techniques in remote sensing for improved image classification. Journal of Geomatics, 11(1), 21.
- Swain, P.H. and Davis, S.M. (1978). Remote Sensing: The Quantitative Approach, McGraw-Hill International Book Co., New York.

14.12 ANSWERS

SAQ I

a) Image classification can be defined as a process of assigning land cover classes/themes to pixels in an image.

b) Training is the process of generating spectral signature of each class.

c) Signature evaluation is a critical step in supervised image classification, as it ensures that the selected training samples accurately represent land cover classes. It involves assessing statistical separability, spectral consistency, and classification performance of training samples before applying any classification algorithm.

d) There are many ways of signature evaluation such as spectral signature analysis, statistical separability analysis, feature space visualisation, signature refinement and optimisation and accuracy prediction and preliminary classification.

SAQ II

a) Please refer to subsection 14.4.1.

b) Please refer to section 14.4.3.

c) Please refer to section 14.5.

d) Please refer to section 14.5.

Terminal Questions

1. Please refer to section 14.2.1.
2. Please refer to section 14.3.2.
3. Please refer to section 14.3.1.
4. Please refer to section 14.8.
5. Please refer to section 14.6.
6. Please refer to section 14.4.
7. Please refer to the subsection 14.3.3.

ignou
THE PEOPLE'S
UNIVERSITY