Learning Theory of Neural Networks for Satellite Data Analysis

大久保, 彰人
九州大学システム情報科学研究科情報理学専攻

https://doi.org/10.11501/3166849

出版情報：九州大学, 1999, 博士（理学）, 課程博士
バージョン：
権利関係：

Learning Theory of Neural Networks for Satellite Data Analysis

大久保 彰人

(i)

Learning Theory of Neural Networks for Satellite Data Analysis

February 2000

大久保彰人

Abstract

Neural networks provide an excellent tool for analyzing remote sensing data in the fields of environment, agriculture, fishery industry, water resources and etc. So, many researchers of remote sensing have tried to apply the neural network to extract physical amounts such as temperature and moisture, and to classify the state of the surface on the earth, using satellite data. There are many types of neural networks for which various learning techniques have been developed until now. In the field of remote sensing, the backpropagation (BP) method and its variants are frequently used for learning neural networks. Self-organizing learning and Hopfield type of learning are also applied to the construction of neural networks.

In the BP method, a multi-layered neural network is learnt so that output errors of the network are minimized for training data. Therefore, the network learnt has classification ability for training data. Data except for training ones are input into the learnt network and the category of the input data is identified. That is, we can not know what category input data belong to, without inputting data in the network. The self-organizing learning technique does not need training data. The learning process proceeds automatically and input data are clustered in some categories. However, it is difficult to interpret such categories.

This thesis proposes two learning methods for three-layered neural networks based on the concept of domains of recognition to analyze remote sensing data having the form of images. The first piece of research is to design a three-layered neural network with one output unit by adding hidden units successively to simplify the structure of the network. Cone-like domains of recognition are introduced to be able to estimate data except for training ones. Furthermore, we apply this network to estimate soil moisture in the plain and to make a soil moisture map. In the second piece of research, we propose a

learning method of three-layered neural networks based on domains of recognition with a nonlinear type of boundaries. The neural network learnt by this method is applied to land cover classification problems. Data to be classified, which are observed by the Thematic Mapper (TM) and the Synthetic Aperture Radar (SAR), are converted into orthogonal components by the principle component analysis to get high accuracy of classification. We give three kinds of simulation results. The first two simulations are carried out using TM data only. In the last one, TM and SAR data are used. We also make land cover classification maps based on the classification results.

Acknowledgments

I would like to extend my sincere gratitude to my supervisor, Prof. Koichi Niijima for his unfailing help and encouragement, as well as the valuable criticisms he offered. Thanks to his favor, I could continue my research.

I would like to thank Prof. Setsuo Arikawa and Prof. Fumihiro Matsuo for their many valuable and adequate comments.

I would like to thank Prof. Yoshifumi Yasuoka of the University of Tokyo. I would never have been able to complete my work on remote sensing without his support.

I also would like to express my gratitude to the head Motohiro Kato and the busy staffs in Fukuoka Institute of Health and Environmental Sciences.

I appreciate the support from the Department of Informatics. Their pleasant advice helped promote my research.

Last but not least, my thanks go to my family for their warmth, support and encouragements, without which I would never have been able to achieve what I was able to achieve.

Contents

Chapter 1

Introduction

The usage of remote sensing that has artificial satellites as a platform is extending in such various fields as environment, agriculture, fishery industry and water resources, by the following reasons:

1. It makes possible the observation of wide areas on the earth at once,
2. Since the satellites go round the earth periodically and fast, we can repeat the observations easily and almost without time delay,
3. The data obtained by remote sensing observations are multi-channel.

In the analysis of satellite data, it is very important to extract physical amounts like temperature and moisture, and to classify the state of the surface on the earth, based on the data of the reflectance, the scattering and the radiation of the electromagnetic waves.

So far, many statistical approaches such as the multiple regression analysis and the maximum likelihood method have been used for the analysis [6, 10, 17, 18, 20, 23, 24]. However, such statistical methods require explanatory variables in addition to the satellite data, and the assumption that the population of the data has a normal distribution.

Recently, various methods using neural networks have been developed for the analysis of remote sensing data. In the papers [1, 2, 3, 8, 13, 23, 27, 28, 29, 30], the backpropagation

(BP) method [26] has been employed to classify satellite images. The paper [5] applied a Hopfield model for feature tracking and recognition from satellite images. The papers [9, 30] used a self-organizing neural network for category classification. Classification by the BP method is based on a multi-layered neural networks learnt by using training data. Therefore, the training data can be classified certainly, but it is not known until unknown data are input in the trained neural network what category they belong to. Self-organizing neural networks are a clustering machine which does not need training data. We can cluster satellite data with a criterion of the network, however, it is difficult to interpret the clustered data.

This thesis proposes two learning methods for three-layered neural networks to analyze remote sensing data having a form of images. As the first research subject, we design a three-layered neural network with one output unit by adding hidden units successively to simplify the structure of the network. The concept of cone-like domains of recognition is introduced to be possible the estimation of unknown data. Furthermore, we apply this network to estimate soil moisture in the plain and to make a soil moisture map. As the second research subject, we propose a learning method of three-layered neural networks based on domains of recognition with a nonlinear type of boundaries. The network learnt by this method is applied to land cover classification problems.

We now describe each chapter of this thesis in more detail.

Chapter 2 is a survey of the fundamentals of remote sensing to be needed in our analysis.

In Chapter 3, we develop a learning method of three-layered neural networks with one output unit. Between a hidden layer and the output unit, a minimization learning of output errors by adding hidden units is adopted to determine connection weights in the network [14, 15, 19]. The weights connecting the input and hidden layers are learnt based

on cone-like domains of recognition derived under the condition that the output at the newly added hidden unit is close to -1 or 1 for training data.

Chapter 4 is devoted to an application of the learning method proposed in Chapter 3 to soil moisture estimation [19] in the plain. The neural network is learnt using the pair of training data observed by the Synthetic Aperture Radar (SAR) installed in JERS-1 and ERS-2, and soil moisture data gathered in the ground truth. The learning of the network proceeds by adding hidden units successively. Finishing the learning, cone-like domains of recognition are obtained and it is checked which domain SAR data not used in training are contained in. Thus, we can estimate soil moisture at all positions in the plain.

In Chapter 5, we present one more learning method for three-layered neural networks [11, 12, 21, 22]. In the first, we derive domains of recognition under categorized and supervised conditions, without imposing any restriction on hidden layer outputs for training data. Furthermore, it is proved that any pattern in the domain is recognized as a training pattern included in the domain. It is also shown that these domains are mutually disjoint per the categorized and supervised conditions. This means that such domains represent categories for classification. Next, we determine connection weights and thresholds of the network so as to enlarge the domains of recognition in the input space. Since the boundaries of the domain take a complicated form, the region, which is a mapping of the domain into the hidden space, is used to make large the domain of recognition. Using the shape of the region, we derive a cost function to be minimized. A minimizing process for the cost function gives our learning algorithm of the network.

We apply in Chapter 6 our learning method proposed in Chapter 5 to land cover classification problems. Data to be classified are observed by the Thematic Mapper (TM) and SAR. Such data are converted into orthogonal components by the principle component analysis to realize high accuracy of classification. We give three kinds of simulation results

[21, 22]. The first two simulations are carried out using TM data only. In the last one, TM and SAR data will be used. We also make land cover classification maps based on the classification results.

Finally, we present the general conclusions of this thesis in Chapter 7.

Chapter 2

Fundamentals of Remote Sensing

The remote sensing data are useful for solving the environmental problems such as global warming, ozone layer depletion, tropical deforestation, desertification and El Nino phenomena.

In the observation by the artificial satellites, multiband data are usually obtained. In 1972, Landsat satellite was first launched. After that, various artificial satellites with many kinds of sensors were launched. The data observed by these satellites take the form of digital numbers so as to be easily processed by the computer and moreover, these digital numbers are converted into a form of images.

2.1 The Principle of Remote Sensing

Remote sensing is a technology which identifies objects and measures their characteristics without any contact away from the earth. The principle of remote sensing is based on the fact that all objects have peculiar characteristics of reflectance and radiation for different electromagnetic waves. Fig. 2.1 represents the range of electromagnetic waves which are commonly used in remote sensing [25]. In our analysis, we use visible rays, near infrared rays, infrared rays and microwaves. By various sensors installed in the earth observation satellites, we can obtain multiband data.

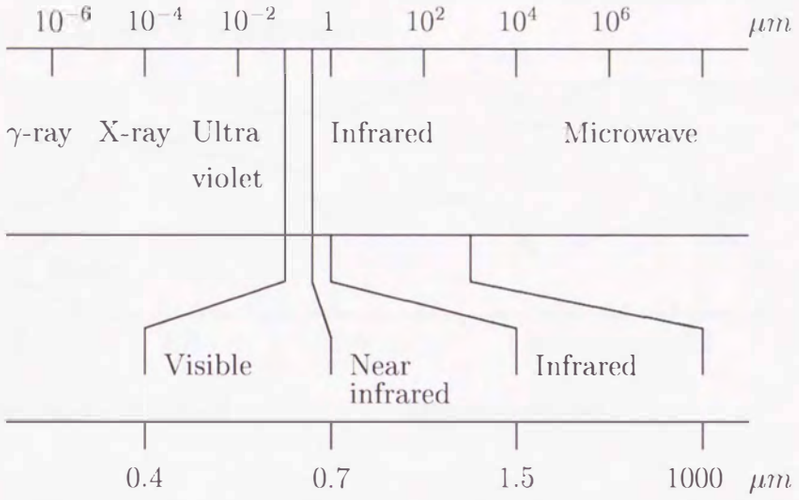


Fig. 2.1: Spectral bands of electromagnetic waves used in remote sensing

Fig. 2.2 shows the reflectance of objects corresponding to various wavelengths [25]. For example, the electromagnetic wave in the near infrared range is absorbed in the water area, and its reflectance becomes small. On the contrary, the vegetation has a larger percentage of reflectance in the near infrared range. These characteristics of various reflectances are used in the analysis of remote sensing.

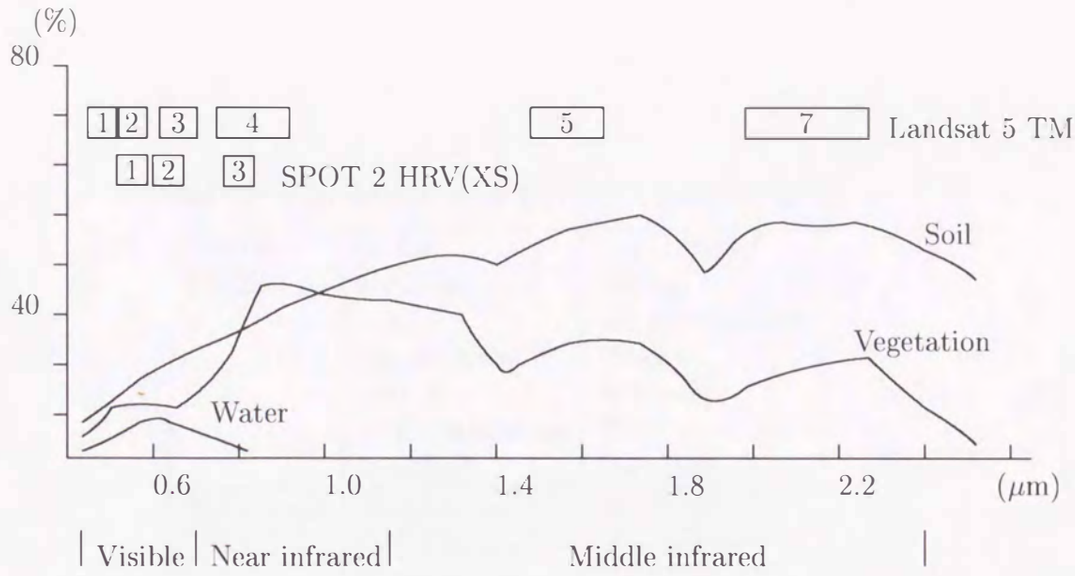


Fig. 2.2: Spectral reflectance characteristics of soil, vegetation and water in the visible and near-to-mid infrared range

2.2 Earth Observation Satellites and Sensors

As earth observation satellites in the visible and infrared regions, we have Landsat 5 and SPOT 2 satellites. The Landsat satellite was the first designed to provide near global coverage of the earth's surface. It has three imaging instruments which are the Return Beam Vidicon, the Multispectral Scanner and TM. In our analysis, we use TM data. TM sensor is a mechanical scanning device. TM sensor covers seven wavelength bands as shown in Table 2.1 [25].

SPOT 2 satellite carries two imaging devices refered to as High Resolution Visible Imaging System (HRV). In our analysis, we use three bands of data in the multispectral mode in the HRV. HRV sensor covers four wavelength bands as shown in Table 2.2 [25].

Next, we describe two satellites JERS-1 and ERS-2 which have SAR and the Active Microwave Instrument (AMI), respectively.

JERS-1 satellite has two imaging instruments: one is an optical sensor, and the other

Table 2.1: Characteristics of Landsat 5 and TM sensor

Satellite	Items	Performance
Landsat 5	altitude	705 km
	orbit	sun synchronous
	repeat cycle	16 days
	period	98.9 min
	orbit inclination	99 °
	launched	Mar 1984
Instrument	Spectral bands	Resolution
TM	0.45-0.52 μm	30 m × 30 m
	0.52-0.60 μm	30 m × 30 m
	0.63-0.69 μm	30 m × 30 m
	0.76-0.90 μm	30 m × 30 m
	1.55-1.75 μm	30 m × 30 m
	10.4-12.5 μm	120 m × 120 m
	2.08-2.35 μm	30 m × 30 m

Table 2.2: Characteristics of SPOT 2 and HRV sensor

Satellite	Items	Performance
SPOT 2	altitude	832 km
	orbit	sun synchronous
	repeat cycle	26 days
	period	101 min
	orbit inclination	99 °
	launched	Jan 1990
Instruments	Spectral bands	Resolution
HRV(XS)	0.50-0.59 μm	20 m × 20 m
	0.60-0.68 μm	20 m × 20 m
	0.79-0.89 μm	20 m × 20 m
HRV(P)	0.51-0.73 μm	10 m × 10 m

Table 2.3: Characteristics of JERS-1 and SAR sensor

Satellite	Items	Performance
JERS-1	altitude	568 km
	orbit	sun synchronous
	repeat cycle	44 days
	period	96 min
	orbit inclination	98 °
	launched	Feb 1992
Instrument	Items	Performance
SAR	frequency	1.275 GHz(L)
	wavelength	23.5 cm
	polarization	HH
	incidence angle	35 °
	swath width	75 km
	resolution	18 m × 18 m

Table 2.4: Characteristics of ERS-2 and AMI sensor

Satellite	Items	Performance
ERS-2	altitude	785 km
	orbit	sun synchronous
	repeat cycle	44 days
	period	96 min
	orbit inclination	98 °
	launched	July 1991
Instrument	Items	Performance
AMI	frequency	5.30 GHz(C)
	wavelength	5.7 cm
	polarization	VV
	incidence angle	23 °
	swath width	100 km
	resolution	30 m × 30 m

an imaging radar. The characteristics of the radar are shown in Table 2.3 [25]. In our analysis, we use SAR data. ERS-2 satellite has the same type of imaging radar as in JERS-1. The characteristics of the radar are shown in Table 2.4 [25]. In our analysis, we also use AMI data.

2.3 Data Specification of TM and HRV

The digital values of brightness level are provided by floppy disks, magnetic tapes and CD-ROM disks. We have two data formats. The Band Sequential (BSQ) format of each band is separately arranged. The Band Interlieved by Line (BIL) format line data are arranged in the order of band number, and repeated with respect to the number (Fig. 2.3).

Following such data formats, we can transform the digital values of Landsat 5 TM and SPOT 2 HRV into image data. For example, Fig. 2.4 shows images for the digital values of Landsat 5 TM having 7 wavelength bands. Fig. 2.5 is a Landsat 5 TM false color image composited using 3 bands among 7 bands.

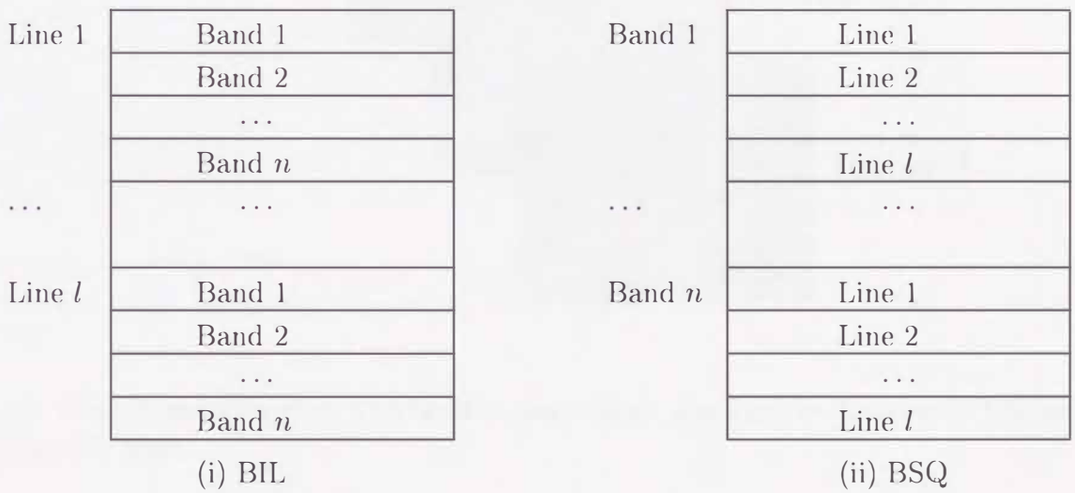


Fig. 2.3: Digital formats of BIL and BSQ

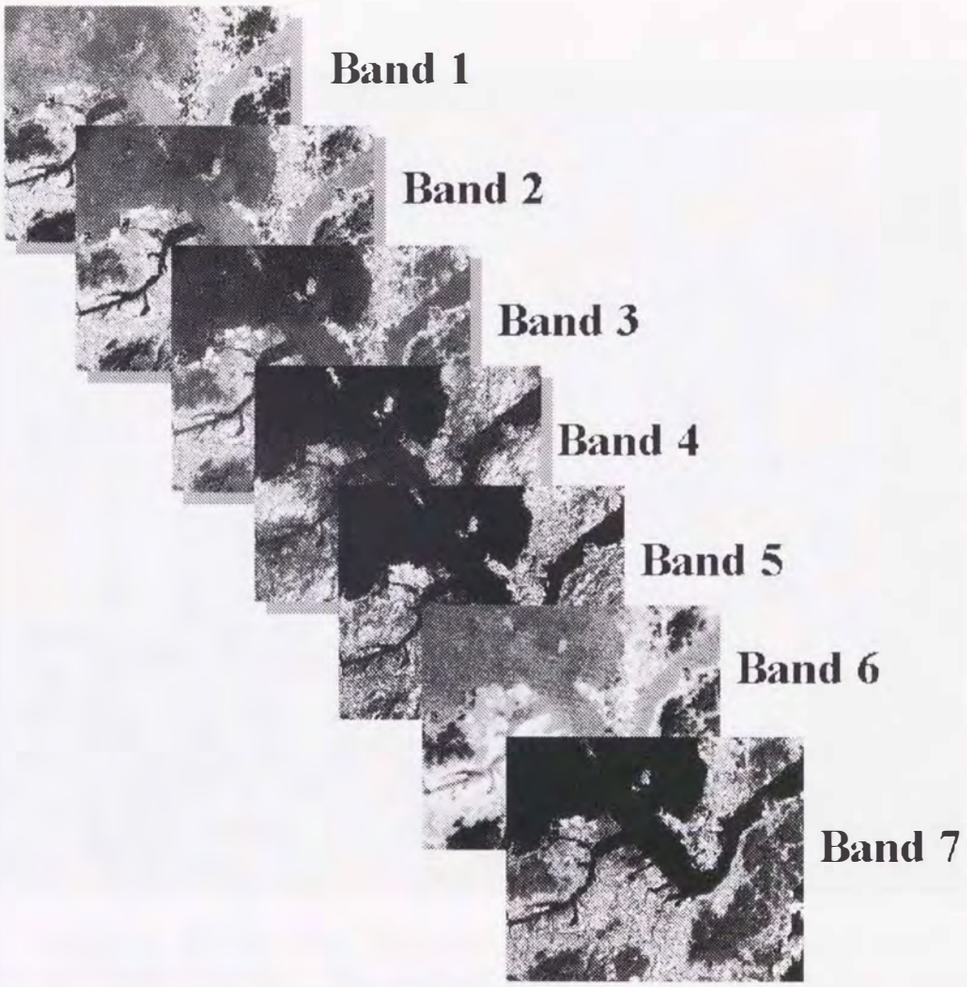


Fig. 2.4: Seven band images of Kitakyushu area, Japan, observed by Landsat 5 TM on September 20, 1990

Fig. 2.5: Landsat 5 TM false color composite image by displaying band 5 as red, band 4 as green and band 3 as blue. The image is enhanced with a stretch from histogram equalization.

2.4 Data Specification of SAR and AMI

In the microwave region, there are active and passive type of sensors. SAR and AMI sensors are of active type. These sensors receive the backscattering which is reflected from the transmitted microwave. The backscattering data are usually gathered using the technique of side looking radar, as illustrated in Fig. 2.6 [25].

The microwave radar has a geometric distortion or shadow depending on the effect of terrain relief, as shown in Fig. 2.7. So, we exclude the mountain area in the land cover classification in Chapter 6.

The digital formats of SAR and AMI consist of 2 bytes data as shown in Fig. 2.8. Following these formats, we can transform the digital values of SAR and AMI into 2 bytes data. However, since such data exceed the range of 0 to 255 (Fig. 2.9), we convert them into 8 bits data to be visible as an image with a gray scale (Fig. 2.10).

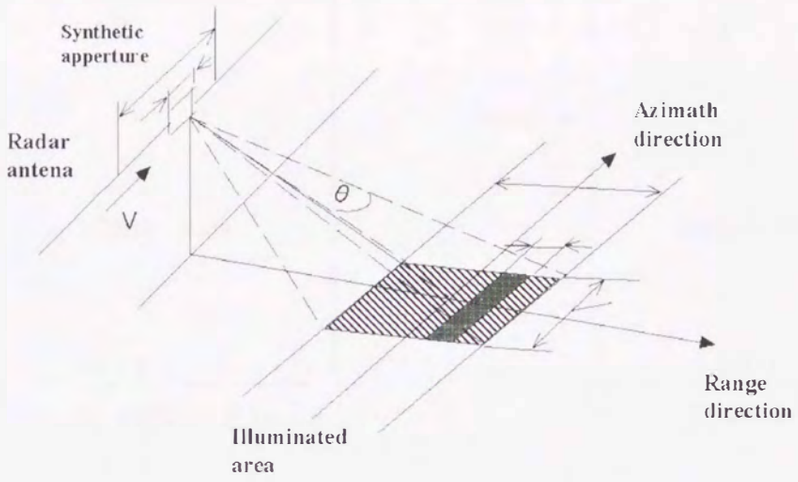


Fig. 2.6: Principle of SAR as a side looking radar

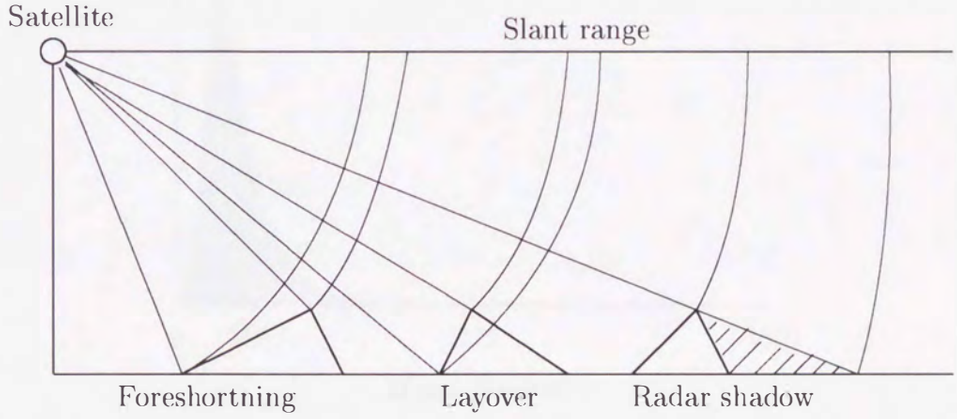


Fig. 2.7: Geometry of radar image

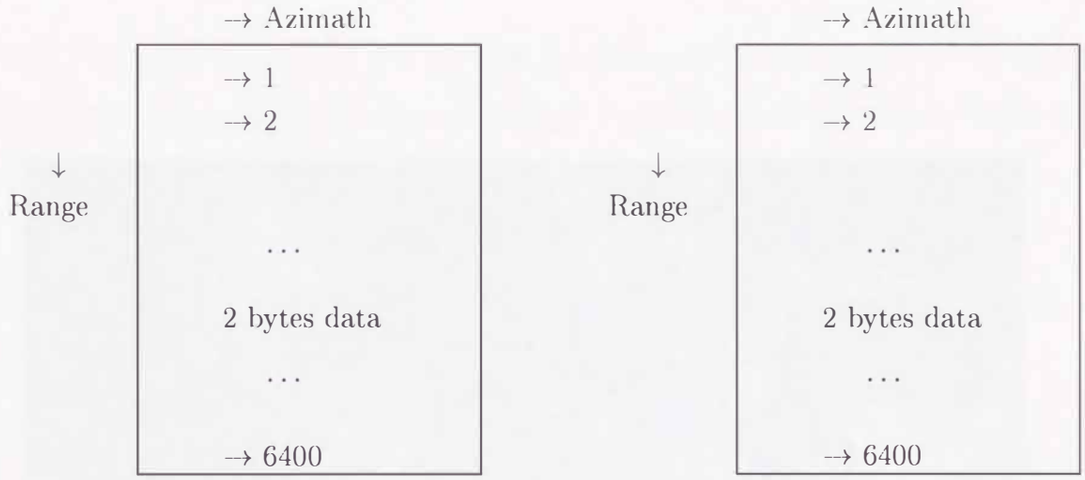


Fig. 2.8: Digital formats of SAR and AMI

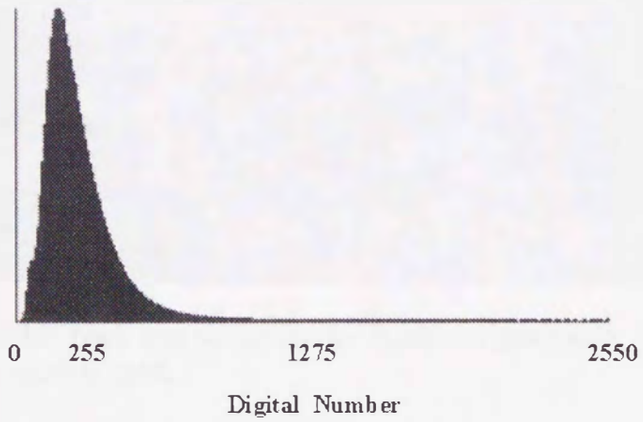


Fig. 2.9: Histogram of 2 bytes digital numbers

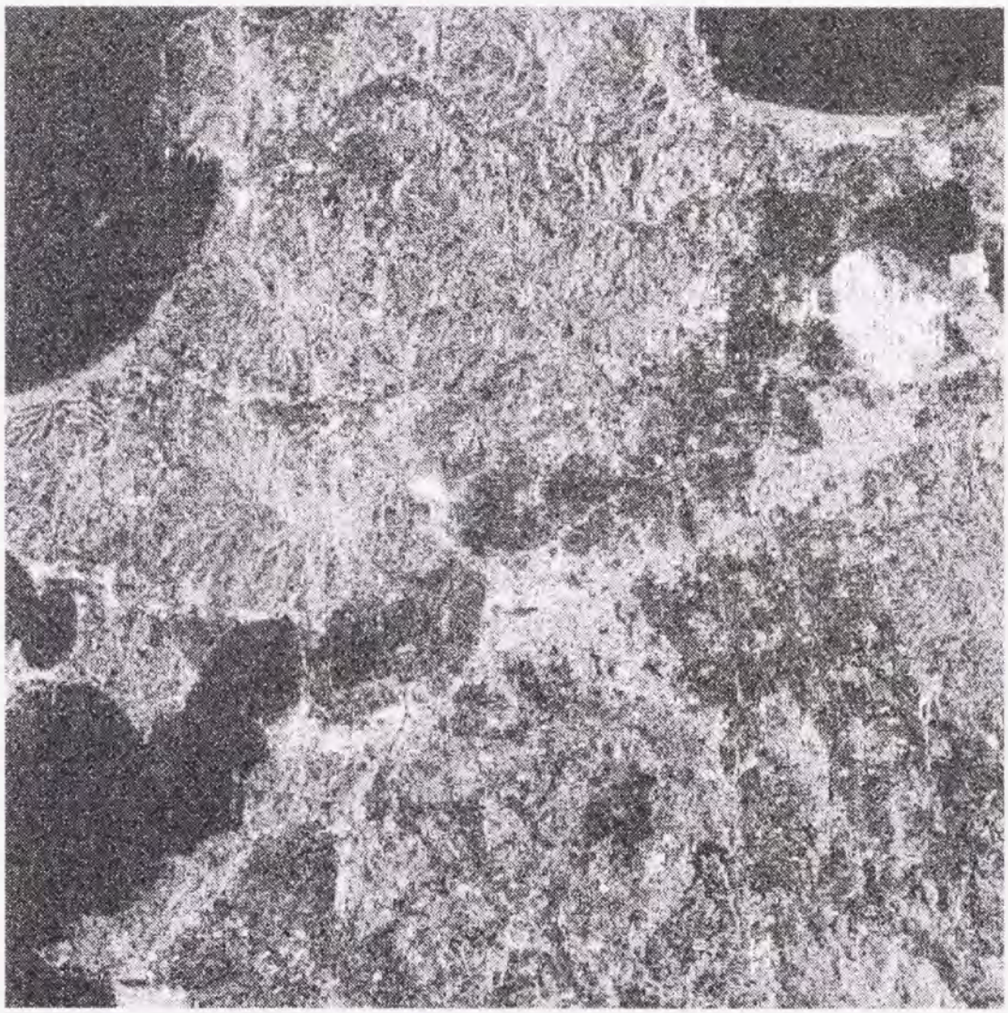


Fig. 2.10: JERS-1 SAR back scatter image for Itoshima peninsula, Japan, with a gray scale, acquired on August 9, 1995

2.5 Geometric Correction

There are two techniques that can be used to correct the various types of geometric distortion present in digital image data. We use one of them, whose approach depends upon establishing mathematical relationships between the addresses of pixels in an image and the corresponding coordinates of those points on the ground.

Suppose that two coordinate systems are related via a pair of affine functions f and g so that

$$u = f(x, y), \quad v = g(x, y).$$

Let (x_i, y_i), $i = 1, 2, \cdots, n$, be ground control points on a map, and (u_i, v_i) corresponding addresses of pixels in an image. Using the least square method, that is,

$$\sum_{i=1}^{n} \Big((u_i - f(x_i, y_i))^2 + (v_i - g(x_i, y_i))^2 \Big) \to \min,$$

we determine the coefficients of affine functions.

By this transform, we can compute (u, v) for any point (x, y) on the map. However, the computed (u, v) does not always correspond to an address of a pixel in the image as shown in Fig. 2.11. So, we interpolate (u, v) using several values of neighboring pixels by approximation methods such as nearest neighbor resampling, bilinear interpolation and cubic convolution interpolation [25]. In Fig. 2.12, the left hand image was transferred to the right hand image by nearest neighbor resampling.

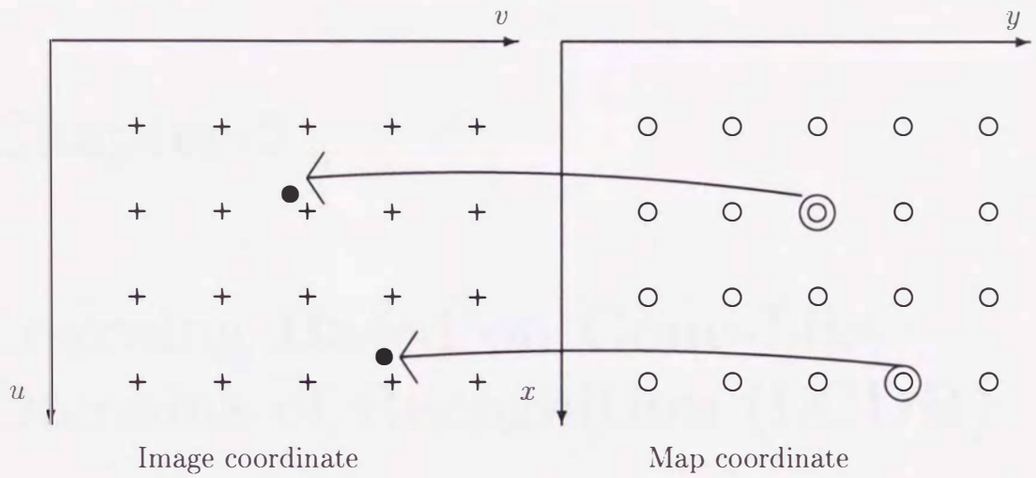


Fig. 2.11: Coordinate conversion for resampling

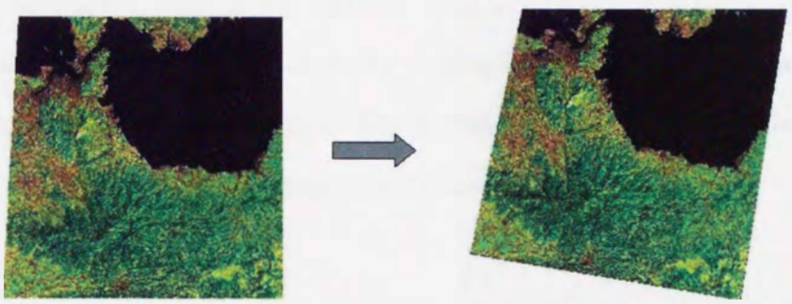


Fig. 2.12: Original and geometric correction images by nearest neighbor resampling

Chapter 3

Learning Based on Cone-Like Domains of Recognition (LCDR)

We propose a learning method of three-layered neural networks based on a successive addition of hidden units and cone-like domains of recognition in the input space. Our approach consists of two learning methods. One is related to a minimization of output errors for a training set, such as BP method. The minimization learning of output errors is done by adding hidden units successively to simplify the structure of the network. The other concerns a maximization of cone-like domains of recognition derived by imposing firing conditions on hidden layer outputs for training input data.

In Section 3.1, we describe a three-layered neural network with one output unit. Section 3.2 is devoted to discuss a learning method by adding hidden units successively. We introduce in Section 3.3 cone-like domains of recognition in the input space. Finally, we give a learning method based on the cone-like domains of recognition, which is described in Section 3.4.

3.1 Three-Layered Neural Network with One Output Unit

We consider a three-layered neural network with one output unit:

$$y = g\left(\sum_{i=1}^{h} w_i f\left(\sum_{k=1}^{n} v_{ik} x_k - \theta_i\right)\right) \tag{3.1}$$

with n input nodes and h hidden units, where x_k are inputs, v_{ik} connection weights between the input and output layers, θ_i indicate thresholds, w_i denote weights connecting the hidden and output layers, and y is an output. The functions $f(t)$ and $g(t)$ are sigmoid functions given by

$$f(t) = \frac{1 - e^{-t}}{1 + e^{-t}}, \qquad g(t) = \frac{1}{1 + e^{-t}}. \tag{3.2}$$

These functions are shown in Fig. 3.1 and Fig. 3.2, respectively.

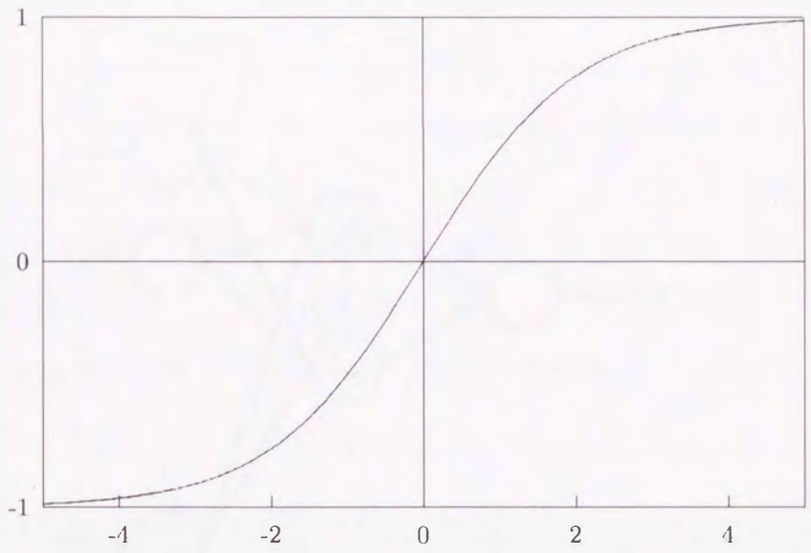


Fig. 3.1: Sigmoid function $f(t)$

We introduce notations $x = {}^t(x_1, x_2, \cdots, x_n)$ and $V_i = {}^t(v_{i1}, v_{i2}, \cdots, v_{in})$, and define $\varphi_i(x)$ by

$$\varphi_i(x) = f(V_i \cdot x - \theta_i),$$

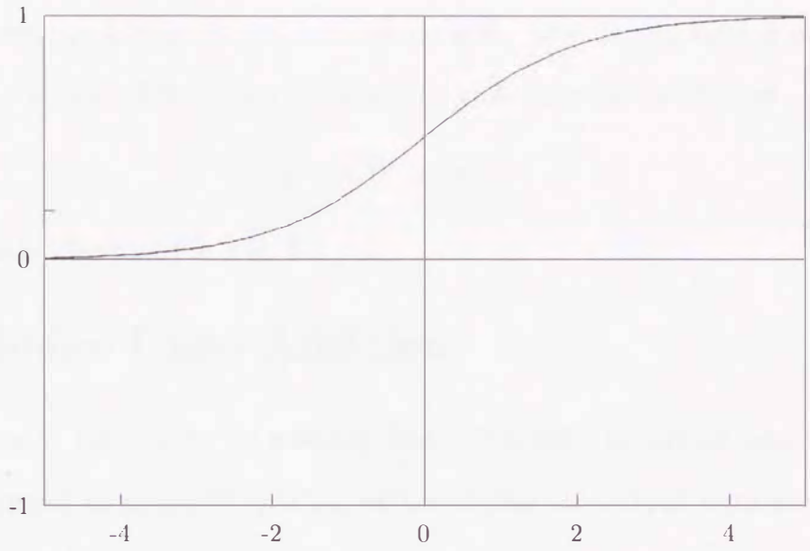


Fig. 3.2: Sigmoid function $g(t)$

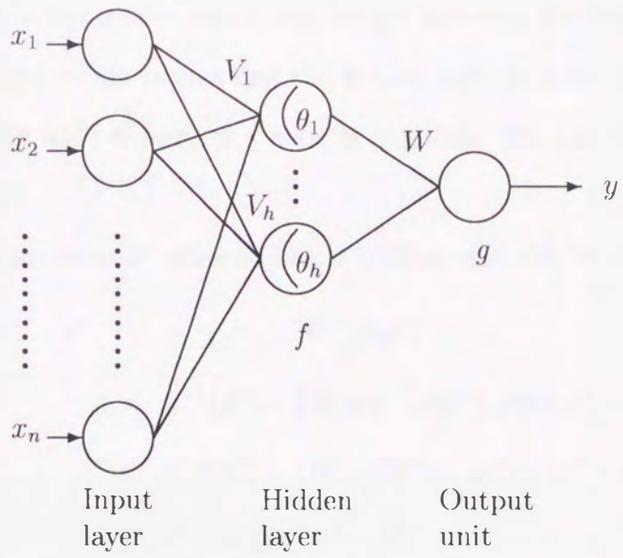


Fig. 3.3: Three-layered neural network with one output unit

where ^{t} denotes the transpose symbol, and $\cdot$ the inner product symbol. The function $\varphi_i(x)$ represents an output at the i-th hidden unit. Introducing further notations $W = {}^t(w_1, w_2, \cdots, w_h)$ and $\varphi(x) = {}^t(\varphi_1(x), \varphi_2(x), \cdots, \varphi_h(x))$, we write (3.1) as

$$y = g(W \cdot \varphi(x)).$$

This relation is illustrated in Fig. 3.3.

3.2 Hidden Units Addition

Let (x^ν, y^ν), $\nu = 1, 2, \cdots, m$, be training data. Although the output error for x^ν and y^ν may be expressed as $y^\nu - g(W \cdot \varphi(x^\nu))$, we now define the output error using the inverse function $g^{-1}(s)$ of $g(t)$ as

$$c^\nu = g^{-1}(y^\nu) - W \cdot \varphi(x^\nu).$$

We assume that V_i, θ_i and W have already been learnt. Adding one more unit to the hidden layer, we define a new connection weight between the hidden unit and the output unit by w, a weight vector connecting the hidden unit with the input layer by v, and the threshold θ in the hidden unit as shown in Fig. 3.4. We put $\tilde{W} = (W, w)$ and $\tilde{\varphi}(x) = (\varphi(x), f(v \cdot x - \theta))$.

The new output error $\tilde{c}^\nu$ after adding a hidden unit can be written as

$$\begin{aligned} \tilde{c}^\nu &= g^{-1}(y^\nu) - \widetilde{W} \cdot \tilde{\varphi}(x^\nu) \\ &= g^{-1}(y^\nu) - (W, w) \cdot (\varphi(x^\nu), f(v \cdot x^\nu - \theta)) \\ &= g^{-1}(y^\nu) - (W \cdot \varphi(x^\nu) + w f(v \cdot x^\nu - \theta)) \\ &= c^\nu - w f(v \cdot x^\nu - \theta). \end{aligned} \tag{3.3}$$

We here calculate the difference L between a squared summation of $\tilde{c}^\nu$ and that of c^ν:

$$L = \sum_{\nu=1}^{m} (\tilde{c}^\nu)^2 - \sum_{\nu=1}^{m} (c^\nu)^2. \tag{3.4}$$

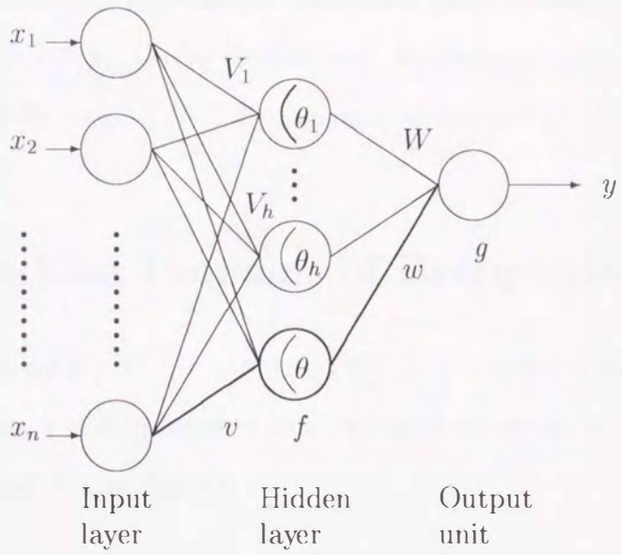


Fig. 3.4: Three-layered neural network after adding one unit in the hidden layer in Fig. 3.3

By (3.3), we have

$$L = \sum_{\nu=1}^{m} (c^\nu - w f(v \cdot x^\nu - \theta))^2 - \sum_{\nu=1}^{m} (c^\nu)^2.$$

An easy calculation yields

$$L = \sum_{\nu=1}^{m} f^2(v \cdot x - \theta) \left(w - \frac{\sum_{\nu=1}^{m} f(v \cdot x^\nu - \theta) c^\nu}{\sum_{\nu=1}^{m} f^2(v \cdot x^\nu - \theta)} \right)^2 - \frac{\left(\sum_{\nu=1}^{m} f(v \cdot x^\nu - \theta) c^\nu\right)^2}{\sum_{\nu=1}^{m} f^2(v \cdot x^\nu - \theta)}. \tag{3.5}$$

This implies that the functional L is minimized when w is chosen as

$$w = \frac{\sum_{\nu=1}^{m} f(v \cdot x^\nu - \theta) c^\nu}{\sum_{\nu=1}^{m} f^2(v \cdot x^\nu - \theta)}.$$

Putting

$$I(v, \theta) = \frac{\left(\sum_{\nu=1}^{m} f(v \cdot x^\nu - \theta) c^\nu\right)^2}{\sum_{\nu=1}^{m} f^2(v \cdot x^\nu - \theta)},$$

it follows from the definition of L and (3.5) that

$$\sum_{\nu=1}^{m} (\tilde{c}^\nu)^2 = \sum_{\nu=1}^{m} (c^\nu)^2 - I(v, \theta). \tag{3.6}$$

It is desirable for $I(v, \theta)$ to be large. There are many methods for determining v and θ so as to maximize $I(v, \theta)$. In the next section, we propose a method for determining such parameters with the help of cone-like domains of recognition in the network.

3.3 Cone-Like Domains of Recognition

We assume that $\varphi_i(x^\nu) \leq -1 + \varepsilon$ or $\varphi_i(x^\nu) \geq 1 - \varepsilon$ holds for V_i, θ_i and W already determined, where ε is a sufficiently small number satisfying $0 < \varepsilon < 1/2$. We define two index sets $I_{\nu,-}$ and $I_{\nu,+}$ as follows:

$$\begin{aligned} I_{\nu,-} &= \{i \mid \varphi_i(x^\nu) \leq -1 + \varepsilon\}, \\ I_{\nu,+} &= \{i \mid \varphi_i(x^\nu) \geq 1 - \varepsilon\}. \end{aligned}$$

Of course, we have $I_{\nu,-} \cup I_{\nu,+} = \{1, 2, \cdots, h\}$.

We first consider a domain of x satisfying

$$\varphi_i(x) \leq \varphi_i(x^\nu), \qquad i \in I_{\nu,-} \tag{3.7}$$

and

$$\varphi_i(x^\nu) \leq \varphi_i(x), \qquad i \in I_{\nu,+}. \tag{3.8}$$

The condition (3.7) implies

$$f(V_i \cdot x - \theta_i) \leq f(V_i \cdot x^\nu - \theta_i), \qquad i \in I_{\nu,-}$$

which is equivalent to

$$V_i \cdot (x - x^\nu) \leq 0, \qquad i \in I_{\nu,-}.$$

Similarly, (3.8) is equivalent to

$$V_i \cdot (x - x^\nu) \geq 0, \qquad i \in I_{\nu,+}. \tag{3.9}$$

Therefore, the domain of x satisfying (3.7) and (3.8) is a cone in the input space as

$$Cone(x^\nu) = \{x \in R^n \mid V_i \cdot (x - x^\nu) \leq 0, \quad i \in I_{\nu,-},$$
$$V_i \cdot (x - x^\nu) \geq 0, \quad i \in I_{\nu,+}\}.$$

Moreover, we define a larger domain than $Cone(x^\nu)$ as

$$D_\rho(x^\nu) = \{x \in R^n \mid V_i \cdot (x - x^\nu) \leq \rho\,|V_i \cdot x^\nu - \theta_i|, \quad i \in I_{\nu,-},$$
$$V_i \cdot (x - x^\nu) \geq -\rho|V_i \cdot x^\nu - \theta_i|, \quad i \in I_{\nu,+}\}, \tag{3.10}$$

where $0 < \rho < 1$. The domains $Cone(x^\nu)$ and $D_\rho(x^\nu)$ are illustrated in Fig. 3.5.

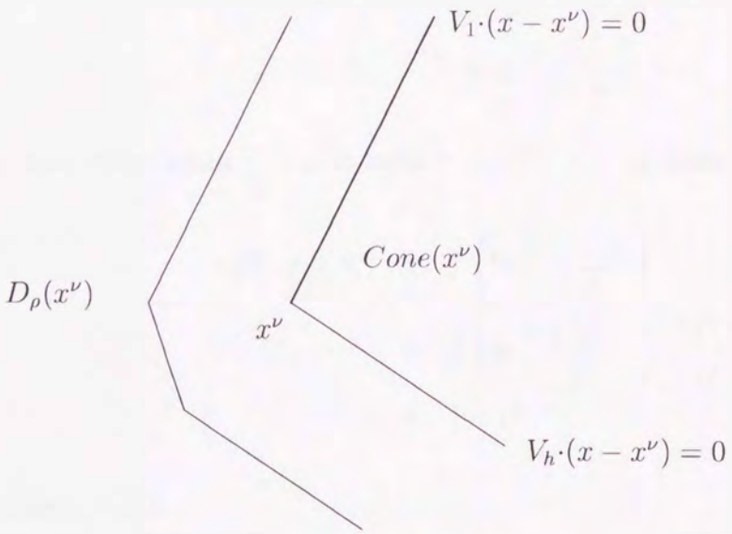


Fig. 3.5: $Cone(x^\nu)$ and $D_\rho(x^\nu)$

We have the following result.

Theorem 3.1. For any x in $D_\rho(x^\nu)$, we have

$$\varphi_i(x) \geq 1 - \varepsilon^{1-\rho}, \qquad i \in I_{\nu,+}, \tag{3.11}$$
$$\varphi_i(x) \leq -1 + \varepsilon^{1-\rho}, \qquad i \in I_{\nu,-}. \tag{3.12}$$

Proof. We choose any x in $D_\rho(x^\nu)$. For $i \in I_{\nu,+}$, we have by the definition of (3.10),

$$\begin{aligned} V_i \cdot x - \theta_i &= V_i \cdot (x - x^\nu) + V_i \cdot x^\nu - \theta_i \\ &\geq -\rho(V_i \cdot x^\nu - \theta_i) + V_i \cdot x^\nu - \theta_i \\ &= (1-\rho)(V_i \cdot x^\nu - \theta_i) \\ &\geq (1-\rho) \ln \frac{2-\varepsilon}{\varepsilon}, \end{aligned}$$

where we used $f(V_i \cdot x^\nu - \theta_i) \geq 1 - \varepsilon$ and $f^{-1}(s) = \ln((1+s)/(1-s))$ in the last line. This inequality and the monotonicity of $f(t)$ lead us to

$$f(V_i \cdot x - \theta_i) \geq f\left((1-\rho) \ln \frac{2-\varepsilon}{\varepsilon}\right). \tag{3.13}$$

Applying here the inequality

$$(1-\rho) \ln \frac{2-\varepsilon}{\varepsilon} \geq \ln \frac{2-\varepsilon^{1-\rho}}{\varepsilon^{1-\rho}} \tag{3.14}$$

to (3.13), and using again $f^{-1}(s) = \ln((1+s)/(1-s))$, we have

$$\begin{aligned} f(V_i \cdot x - \theta_i) &\geq f\left(\ln \frac{2-\varepsilon^{1-\rho}}{\varepsilon^{1-\rho}}\right) \\ &= f\left(\ln \frac{1+(1-\varepsilon^{1-\rho})}{1-(1-\varepsilon^{1-\rho})}\right) \\ &= 1 - \varepsilon^{1-\rho} \end{aligned}$$

which implies (3.11).

On the other hand, we have for $i \in I_{\nu,-}$,

$$\begin{aligned} V_i \cdot x - \theta_i &= V_i \cdot (x - x^\nu) + V_i \cdot x^\nu - \theta_i \\ &\leq -\rho(V_i \cdot x^\nu - \theta_i) + V_i \cdot x^\nu - \theta_i \\ &= (1-\rho)(V_i \cdot x^\nu - \theta_i) \\ &\leq -(1-\rho) \ln \frac{2-\varepsilon}{\varepsilon}. \end{aligned}$$

This inequality and the monotonicity of $f(t)$ give

$$f(V_i \cdot x - \theta_i) \leq f\left(-(1-\rho)\ln\frac{2-\varepsilon}{\varepsilon}\right) \tag{3.15}$$

Applying (3.14) again to (3.15), we have

$$\begin{aligned} f(V_i \cdot x - \theta_i) &\leq f\left(\ln\frac{\varepsilon^{1-\rho}}{2-\varepsilon^{1-\rho}}\right) \\ &= f\left(\ln\frac{1+(-1+\varepsilon^{1-\rho})}{1-(-1+\varepsilon^{1-\rho})}\right) \\ &= -1+\varepsilon^{1-\rho} \end{aligned}$$

which proves (3.12).

This theorem implies that the output vector $\varphi(x) = (\varphi_1(x), \varphi_2(x), \cdots, \varphi_h(x))$ for any x in $D_\rho(x^\nu)$ is almost the same as $\varphi(x^\nu)$, that is, $\varphi(x)$ can be recognized as $\varphi(x^\nu)$.

For the new weight vector v and the threshold θ, we define

$$\varphi_{h+1}(x) = f(v \cdot x - \theta)$$

and assume that

$$\varphi_{h+1}(x^\nu) \leq -1+\varepsilon \tag{3.16}$$

or

$$\varphi_{h+1}(x^\nu) \geq 1-\varepsilon \tag{3.17}$$

holds. Of course, we can rewrite (3.16) and (3.17) as

$$v \cdot x^\nu - \theta \leq -\ln\frac{2-\varepsilon}{\varepsilon} \tag{3.18}$$

and

$$v \cdot x^\nu - \theta \geq \ln\frac{2-\varepsilon}{\varepsilon}, \tag{3.19}$$

respectively.

We define two domains

$$D_\rho^-(x^\nu) = D_\rho(x^\nu) \cap \{x \in R^n \,|\, v \cdot (x - x^\nu) \leq \rho\,|v \cdot x^\nu - \theta|\}$$

and

$$D_\rho^+(x^\nu) = D_\rho(x^\nu) \cap \{x \in R^n \,|\, v \cdot (x - x^\nu) \geq -\rho\,|v \cdot x^\nu - \theta|\},$$

and two index sets

$$I^*_{\nu,-} = \begin{cases} I_{\nu,-} \cup \{h+1\} & \text{if} \quad \varphi_{h+1}(x^\nu) \leq -1+\varepsilon, \\ I_{\nu,-} & \text{if} \quad \varphi_{h+1}(x^\nu) \geq 1-\varepsilon \end{cases}$$

and

$$I^*_{\nu,+} = \begin{cases} I_{\nu,+} & \text{if} \quad \varphi_{h+1}(x^\nu) \leq -1+\varepsilon, \\ I_{\nu,+} \cup \{h+1\} & \text{if} \quad \varphi_{h+1}(x^\nu) \geq 1-\varepsilon. \end{cases}$$

Then, we have the following result.

Corollary 3.2. For any x in $D_\rho^-(x^\nu)$ and in $D_\rho^+(x^\nu)$, we have

$$\varphi_i(x) \;\geq\; 1 - \varepsilon^{1-\rho}, \qquad i \in I^*_{\nu,+}, \tag{3.20}$$

$$\varphi_i(x) \;\leq\; -1 + \varepsilon^{1-\rho}, \qquad i \in I^*_{\nu,-}. \tag{3.21}$$

Proof. It suffices to prove (3.20) and (3.21) for any $x \in D_\rho^+(x^\nu)$. In this case, we have assumed $\varphi_{h+1}(x^\nu) \geq 1-\varepsilon$ and hence, we have $I^*_{\nu,-} = I_{\nu,-}$ and $I^*_{\nu,+} = I_{\nu,+} \cup \{h+1\}$. Since we have already proved (3.20) and (3.21) for $i \in I_{\nu,+}$ and $i \in I_{\nu,-}$, it suffices to prove

$$\varphi_{h+1}(x) \geq 1 - \varepsilon^{1-\rho}, \tag{3.22}$$

for any $x \in D_\rho^+(x^\nu)$, that is,

$$v \cdot x - \theta \geq \ln \frac{2 - \varepsilon^{1-\rho}}{\varepsilon^{1-\rho}}.$$

In the same way as in the proof of Theorem 3.1, we have

$$\begin{aligned} v \cdot x - \theta &= v \cdot (x - x^\nu) + v \cdot x^\nu - \theta \\ &\geq (1 - \rho)(v \cdot x^\nu - \theta) \\ &\geq (1 - \rho) \ln \frac{2 - \varepsilon}{\varepsilon}. \end{aligned}$$

Using (3.14) again, we get

$$v \cdot x - \theta \geq \ln \frac{2 - \varepsilon^{1-\rho}}{\varepsilon^{1-\rho}}$$

which finishes the proof.

This corollary implies that the output vector $(\varphi(x), \varphi_{h+1}(x))$ for any x in $D_\rho^-(x^\nu)$ and in $D_\rho^+(x^\nu)$ is almost the same as $(\varphi(x^\nu), \varphi_{h+1}(x^\nu))$, that is, $(\varphi(x), \varphi_{h+1}(x))$ can be recognized as $(\varphi(x^\nu), \varphi_{h+1}(x^\nu))$.

3.4 Learning Method

From the definition of $D_\rho^-(x^\nu)$ and $D_\rho^+(x^\nu)$, we have the inclusions

$$D_\rho^-(x^\nu) \subset D_\rho(x^\nu)$$

and

$$D_\rho^+(x^\nu) \subset D_\rho(x^\nu).$$

This means that if a hidden unit is added in the hidden layer, the domains of recognition become smaller than before adding. However, the output error becomes smaller before adding hidden units. This is a trade-off problem.

It is desirable for $D_\rho^-(x^\nu)$ and $D_\rho^+(x^\nu)$ to be as large as possible. We may concentrate on $D_\rho^-(x^\nu)$ because the situation is the same for $D_\rho^+(x^\nu)$. Notice here that $D_\rho^-(x^\nu)$ can be

decomposed as $D_\rho^-(x^\nu) = Cone^-(x^\nu) \cup Str^-(x^\nu)$, where

$$Cone^-(x^\nu) = \{x \mid V_i \cdot (x - x^\nu) \leq 0, \quad i \in I_{\nu,-},$$
$$V_i \cdot (x - x^\nu) \geq 0, \quad i \in I_{\nu,+}, \quad v \cdot (x - x^\nu) \leq 0\}$$

and

$$Str^-(x^\nu) = \{x \mid 0 < V_i \cdot (x - x^\nu) \leq \rho \, |V_i \cdot x^\nu - \theta_i|, \quad i \in I_{\nu,-},$$
$$V_i \cdot (x - x^\nu) \geq -\rho \, |V_i \cdot x^\nu - \theta_i|, \quad i \in I_{\nu,+},$$
$$0 < v \cdot (x - x^\nu) \leq \rho \, |v \cdot x^\nu - \theta|\}. \tag{3.23}$$

The domains $D_\rho^-(x^\nu)$, $Cone^-(x^\nu)$ and $Str^-(x^\nu)$ are illustrated in Fig. 3.6.

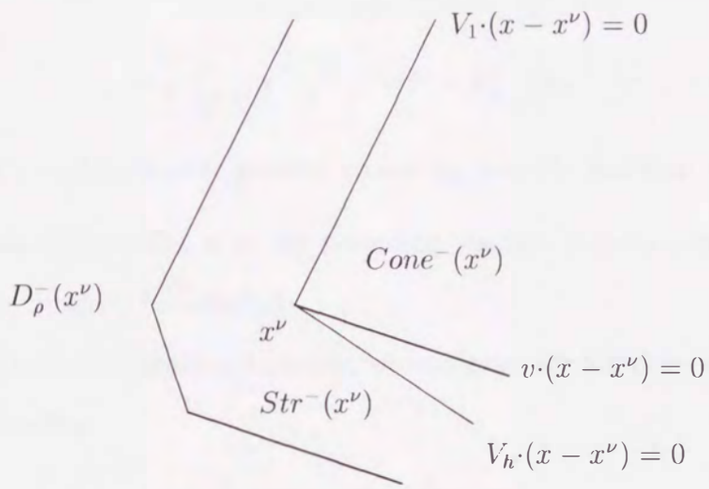


Fig. 3.6: $D_\rho^-(x^\nu)$, $Cone^-(x^\nu)$ and $Str^-(x^\nu)$

To enlarge the domain $D_\rho^-(x^\nu)$, we maximize the width of $Str^-(x^\nu)$ and the angles of $Cone^-(x^\nu)$. The weight vectors V_i and the thresholds θ_i have already been determined and so, it suffices to determine v and θ so as to maximize the width of stripe between the

hyperplanes $v \cdot (x - x^\nu) = 0$ and $v \cdot (x - x^\nu) = \rho|v \cdot x^\nu - \theta|$. The width is given by

$$\frac{\rho|v \cdot x^\nu - \theta_i|}{\|v\|} \tag{3.24}$$

which was derived in [14]. We want to maximize (3.24). However, the quantity (3.24) depends on the index ν and so, $\sum_{\nu=1}^{m}(v \cdot x^\nu - \theta)^2 / \|v\|^2$ is maximized, that is,

$$\frac{\|v\|^2}{\sum_{\nu=1}^{m}(v \cdot x^\nu - \theta)^2} \tag{3.25}$$

is minimized.

Next, we maximize the angle γ at which the hyperplanes $V_j \cdot (x - x^\nu) = 0$ and $v \cdot (x - x^\nu) = 0$ cross. To do so, it suffices to minimize $\cos\gamma = V_j \cdot v / \|V_j\| \|v\|$.

Summarizing the above discussions, we minimize the cost function:

$$\begin{aligned} J(v, \theta) = & \frac{\|v\|^2}{\sum_{\nu=1}^{m}(v \cdot x^\nu - \theta)^2} + C_1 \sum_{j=1}^{h} \frac{V_j \cdot v}{\|V_j\| \|v\|} \\ & + C_2 \sum_{\nu=1}^{m} \left(\ln \frac{2-\varepsilon}{\varepsilon} - v \cdot x^\nu + \theta \right)_+^2 \left(\ln \frac{2-\varepsilon}{\varepsilon} + v \cdot x^\nu - \theta \right)_+^2, \end{aligned} \tag{3.26}$$

where C_1 and C_2 denote penalty constants, and the function z_+^2 is defined by $z_+^2 = z^2$ if $z \geq 0$ and $z_+^2 = 0$ if $z < 0$. By bounding the last penalty term, the condition (3.18) or (3.19) becomes to be satisfied.

In actual computation, however, we minimize the following functional to avoid numerical instability:

$$\begin{aligned} J(U, \beta, \eta) = & \frac{1}{\sum_{\nu=1}^{m}(U \cdot x^\nu - \eta)^2} + C_1 \sum_{j=1}^{h} U_j \cdot U \\ & + C_2 \sum_{\nu=1}^{m} (\alpha - \beta(U \cdot x^\nu - \eta))_+^2 (\alpha + \beta(U \cdot x^\nu - \eta))_+^2 + C_3(\|U\|^2 - 1)^2, \end{aligned} \tag{3.27}$$

where C_3 denotes a penalty constant and we have put

$$U_j = \frac{V_j}{\|V_j\|},$$

$$
\begin{aligned}
U &= \frac{v}{\|v\|}, \\
\eta &= \frac{\theta}{\|v\|}, \\
\beta &= \|v\|
\end{aligned}
$$

and

$$\alpha = \ln \frac{2-\varepsilon}{\varepsilon}.$$

We can minimize (3.27) by using, for example, the gradient methods to obtain U,η and β, and to compute the connection weight vector $v = \beta U$ and the threshold $\theta = \beta\eta$.

We must also minimize $-I(v,\theta) \equiv -I(U,\beta,\eta)$ appeared in Section 3.2. Thus, we minimize finally the following cost function:

$$K(U,\beta,\eta) = J(U,\beta,\eta) - CI(U,\beta,\eta)$$

with a penalty constant C.

Chapter 4

Soil Moisture Estimation by LCDR Method

Soil moisture is one of parameters for estimating the runoff, the evaporation and the transpiration in the water resource problem. Therefore, the soil moisture is an important parameter in the hydrogical process. The relationships between soil moisture and radar measurement were investigated in [4, 7, 24].

In this chapter, using the learning method proposed in Chapter 3, we estimate soil moisture based on the data observed by artificial satellites [19].

4.1 Input Data and Training Data

We apply the learning method based on cone-like domains of recognition proposed in Chapter 3 to estimate soil moisture in the plain and to make a soil moisture map, using AMI, SAR, HRV and TM data.

First, we show how to make the input vector $x = {}^t(\, x_1,\, x_2,\, x_3,\, x_4\,)$ in a three-layered neural network. The first three components x_1, x_2 and x_3 are chosen as

$$x_k = 20\, \log_{10}(I_k) + CF_k, \quad k = 1, 2, 3. \tag{4.1}$$

Here, I_1, I_2 and I_3 denote the backscattering coefficients of electromagnetic waves, which are observed by the artificial satellites ERS-2, JERS-1 and JERS-1, respectively.

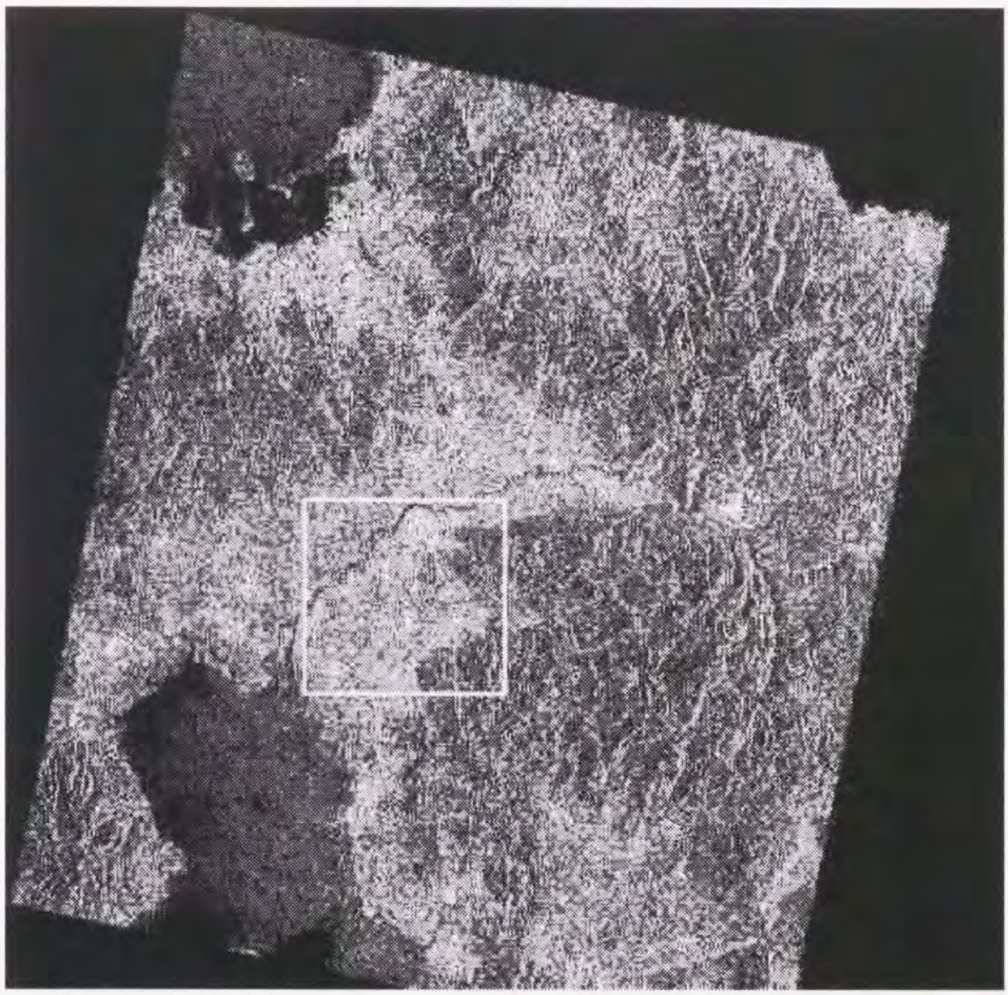


Fig. 4.1: Kyushu island, Japan, observed by ERS-2 AMI on January 17,1997. The object area is Chikushi plain which is shown in the square.

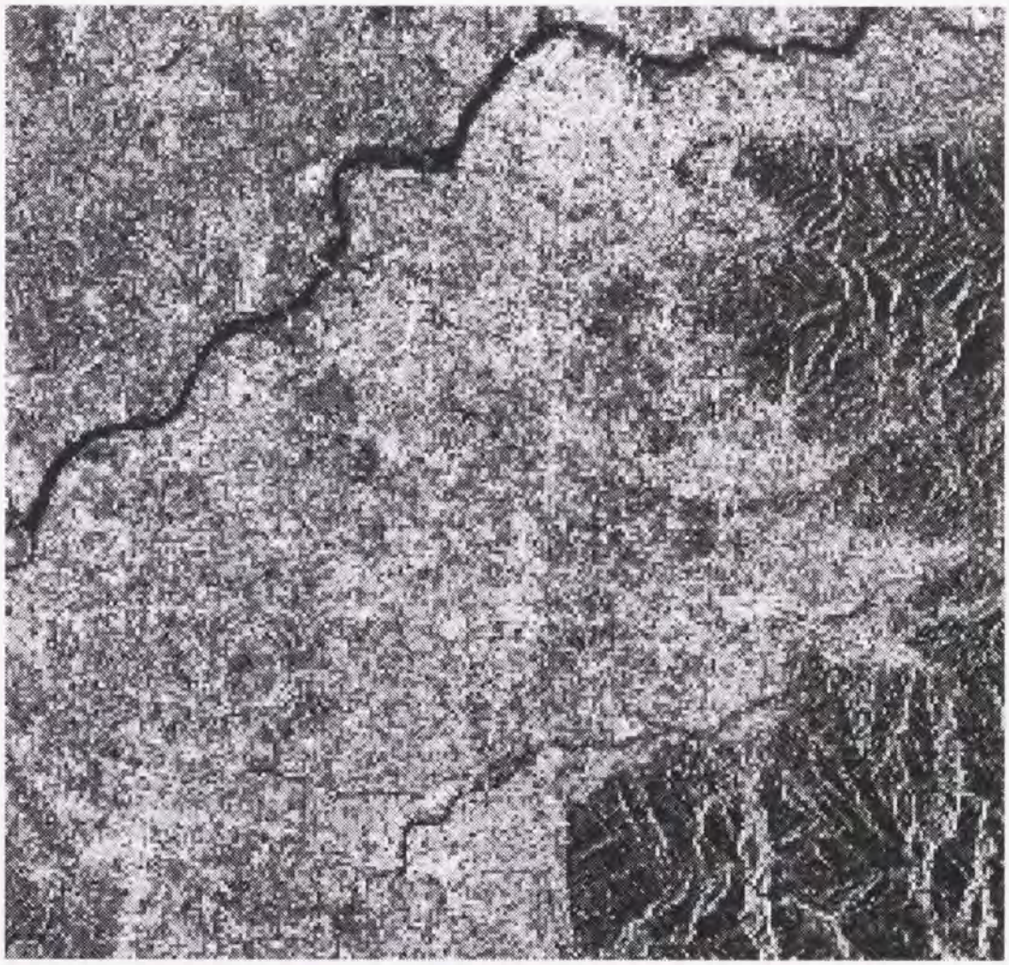


Fig. 4.2: Image of I_1 for Chikushi plain observed by ERS-2 AMI on January 17, 1997

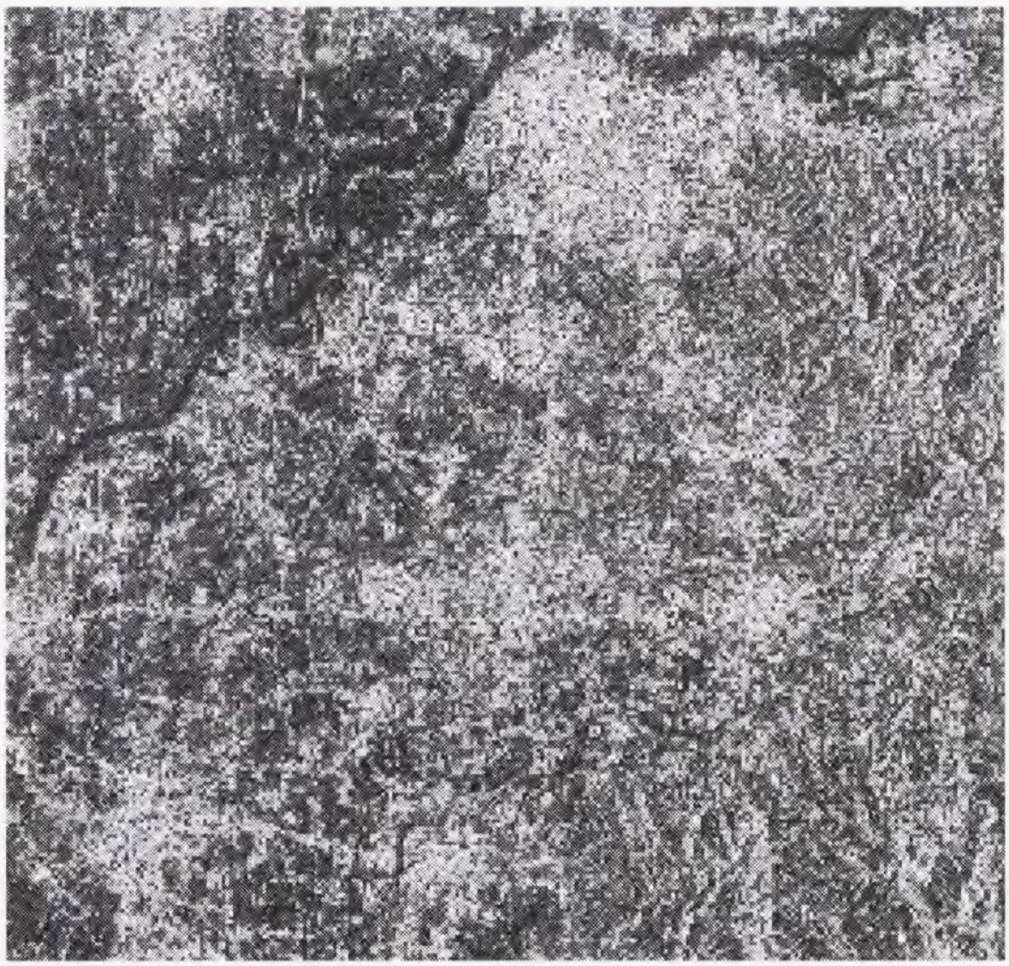


Fig. 4.3: Image of I_2 for Chikushi plain observed by JERS-1 SAR on January 17, 1997

Fig. 4.4: Image of I_3 for Chikushi plain observed by JERS-1 SAR on January 18, 1997

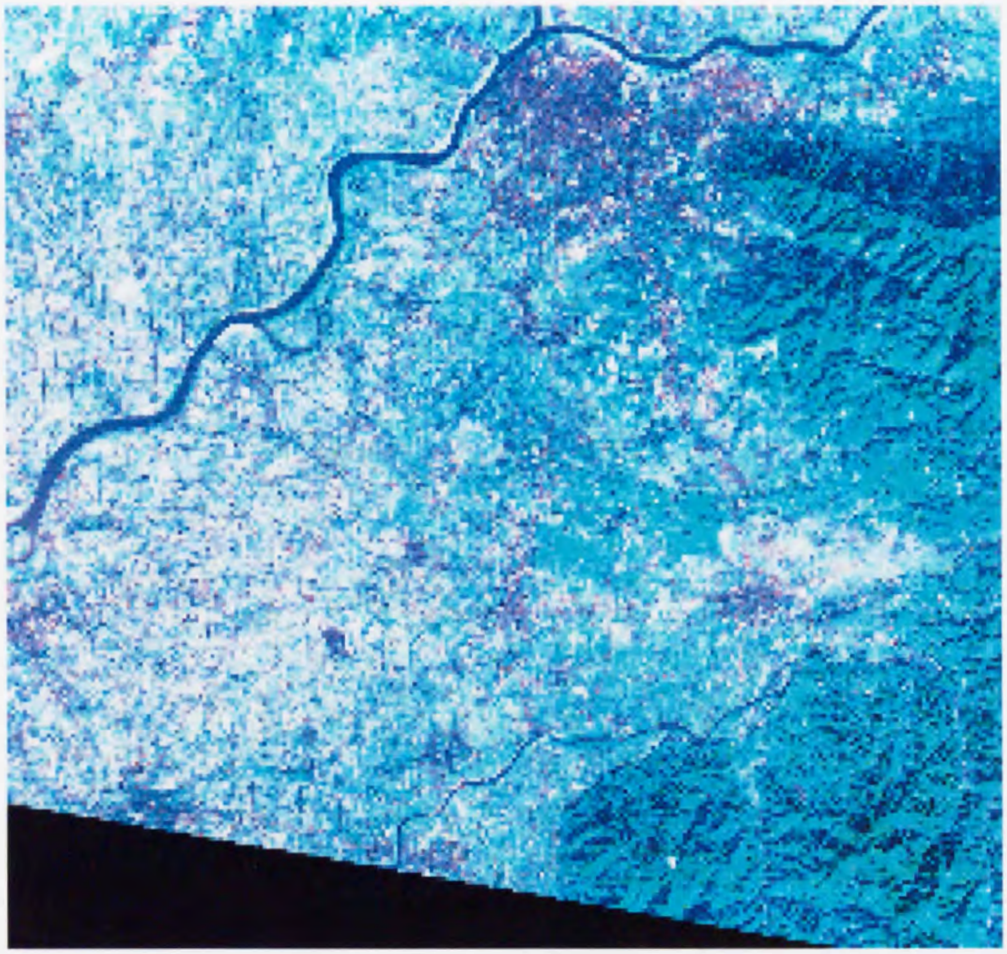


Fig. 4.5: Image for computing NVI for Chikushi plain observed by SPOT 2 HRV(XS) on December 27, 1996. NIR and VIS values are contained in this image.

Fig. 4.6: Image for computing NVI in Chikushi plain observed by Landsat 5 TM on Mar 30, 1994. NIR and VIS values are contained in this image.

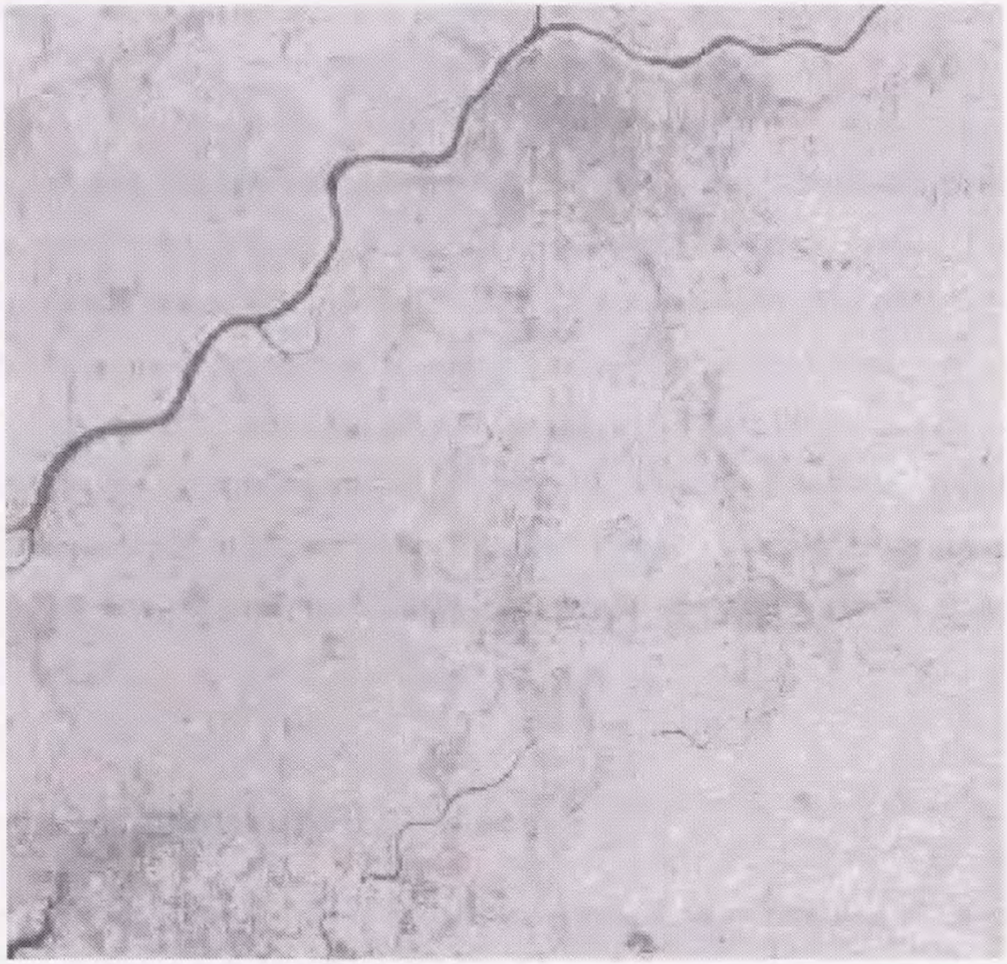


Fig. 4.7: NVI image computed using NIR and VIS values in Fig. 4.5 and Fig. 4.6

The coefficients I_1, I_2 and I_3 are visualized in Figs. 4.2, 4.3 and 4.4, respectively. These were observed in the object area shown in Fig. 4.1. The CF_k denote the calibration coefficients, which are given in advance by National Space Development Agency of Japan (NASDA).

Normal Vegetation Index (NVI) is calculated by

$$NVI = \frac{NIR - VIS}{NIR + VIS}, \tag{4.2}$$

where NIR and VIS reveal the reflectance values in the near-infrared and visible red ranges, respectively. NIR and VIS values are contained in the images shown in Figs. 4.5 and 4.6. These images were observed by SPOT 2 and Landsat 5. The NVI in (4.2) is denoted by x_4 which corresponds to a fourth component of an input vector. Data of the NVI image in Fig. 4.7 were constructed by using the transform $127 \times x_4 + 128$.

Next, we give the training data $(x^\nu, y^\nu), \nu = 1, 2, \cdots, m$. The y^ν indicate soil moisture, which were gathered at 18 positions in the object area shown in Fig. 4.1 by the ground truth. Therefore, we have $m = 18$. The input data x^ν denotes the satellite data corresponding to y^ν.

4.2 Soil Moisture Estimation

Using the learning method for a three-layered neural network described in Chapter 3, we estimate soil moisture in the object area shown in Fig. 4.1. From the construction of data in the previous section, the number n of input nodes is 4 and the number of the training data is 18. In simulations, we chose $\varepsilon = 10^{-10}$, $C_1 = 9.5$, $C_2 = 10.0$ and $C_3 = 1.0$ in the cost function (3.27), and moved the number h of hidden units from 1 to 15. The output values at the hidden layer were as in Table 4.1 in which we wrote "+1" if $f(v \cdot x^\nu - \theta) \geq 1 - \varepsilon$, and "$-1$" if $f(v \cdot x^\nu - \theta) \leq -1 + \varepsilon$. From Table 4.1, we see that the

18 training input data were grouped in 7 classes. Of course, we can obtain the domains of recognition. Table 4.2 says that when the number of hidden units increases, the output errors are decreasing while the number of unclassified pixels increases. According to the statistical regression analysis, the correlation coefficient between the supervised values y^{ν} and their estimates was 0.455 in this problem. Since this value is close to 0.434 in Table 4.2, we adopt a neural network with 5 hidden units. Fig. 4.8 shows a soil moisture map made based on the domains of recognition in this network. In Fig. 4.8, we masked the mountain area because the soil moisture was surveyed only in the plain. The obtained domains of recognition covered 86% in the plain area where the soil moisture could be estimated.

To evaluate our results, we compared with the results obtained by the multiple regression analysis, which are contained in [20]. Only SAR and NVI data were used in our estimation, but in our statistical approach, geographical information and satellite images of high resolution as well as SAR and NVI data were used. Nevertheless, the accuracy of soil moisture estimation was almost the same in both methods.

4.3 Conclusion

We constructed a neural network using the learning method in Chapter 3 for soil moisture estimation. As a result, we could obtain the domains of recognition for classifying the satellite data which enable us to estimate soil moisture values.

Our estimation method is superior to our statistical method [20] from the viewpoint of amounts of explanatory variables used.

Although the increase of the hidden units makes small output errors for training data, the size of domains of recognition becomes small, which is a trade-off problem.

Table 4.1: Output values at the hidden layer for the training data

$\nu \setminus n$	1	2	3	4	5
1	+1	−1	+1	−1	−1
2	+1	−1	+1	+1	+1
3	+1	−1	+1	−1	+1
4	−1	−1	−1	+1	+1
5	−1	−1	−1	−1	+1
6	−1	−1	−1	−1	+1
7	+1	−1	+1	−1	−1
8	−1	−1	−1	−1	−1
9	−1	−1	−1	+1	+1
10	+1	−1	−1	−1	−1
11	+1	−1	+1	−1	−1
12	+1	−1	+1	+1	+1
13	+1	−1	+1	−1	−1
14	+1	−1	+1	+1	+1
15	+1	−1	+1	−1	+1
16	+1	−1	+1	+1	+1
17	+1	−1	+1	−1	−1
18	−1	−1	−1	−1	+1

Table 4.2: Trade-off between output errors and coverage rates ($\rho = 0.9$)

	Number of hidden units		
	5	10	15
Output error	0.991	0.875	0.753
Coverage rate by domains of recognition(%)	86.02	44.46	27.64
Correlation coefficient	0.434	0.530	0.647

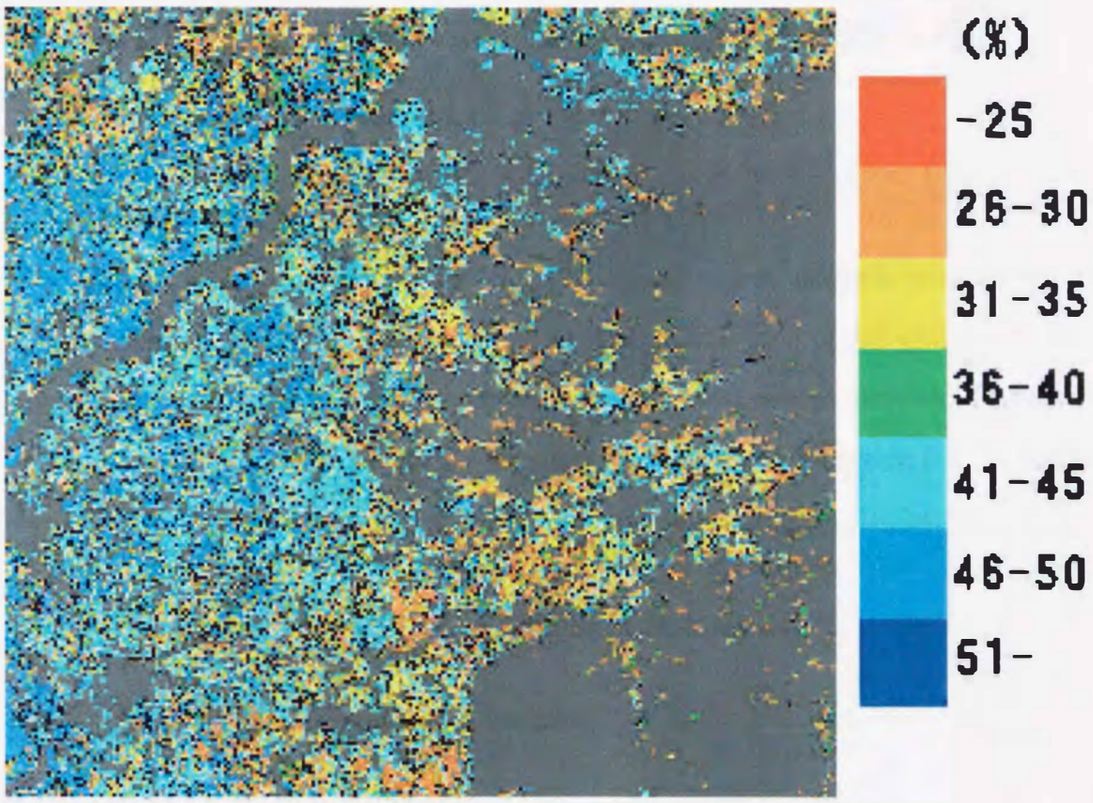


Fig. 4.8: Soil moisture map generated by LCDR method

Chapter 5

Learning Based on Domains of Recognition (LDR)

We propose a new learning method of three-layered neural networks without any restriction of hidden unit outputs, using the concept of domains of recognition in the input space. In Section 5.1, we define a general three-layered neural network. Section 5.2 describes training data. We introduce domains of recognition in Section 5.3. In Section 5.4, we give a learning method based on the domains of recognition [16, 21, 22].

5.1 Three-Layered Neural Network

We consider a three-layered neural network:

$$\begin{aligned} y_i &= g\left(W_i \cdot \psi(x) - \theta_i\right) \\ &\equiv \varphi_i(x), \qquad\qquad i = 1, 2, \cdots, l \end{aligned} \tag{5.1}$$

in which $\psi(x) = {}^t(\psi_1(x), \psi_2(x), \cdots, \psi_h(x))$, where $\psi_j(x) = f\left(V_j \cdot x - \eta_j\right)$, $x = {}^t(x_1, x_2, \cdots, x_n)$ is an input vector, V_j denote weight vectors connecting the input and hidden layers, and θ_i indicate thresholds (Fig. 5.1). The W_i denote connection weight vectors between the hidden and output layers, and y_i are outputs. The functions $f(t)$ and $g(t)$ have already been given in (3.2) in Section 3.1.

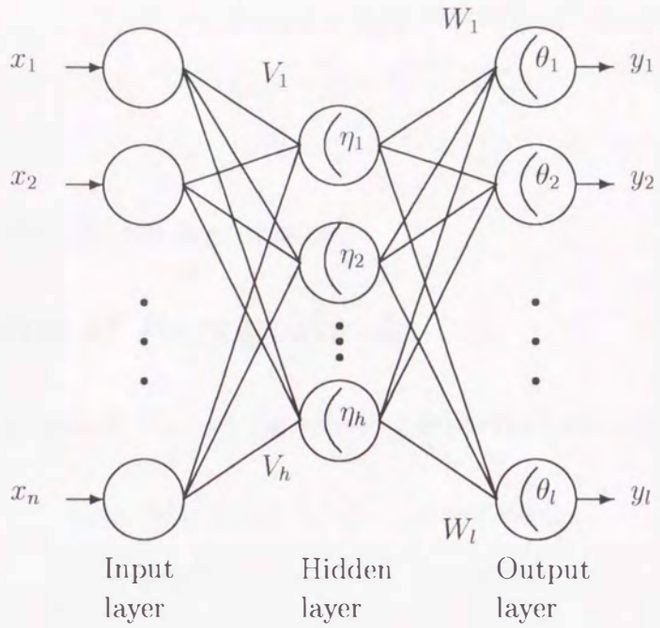


Fig. 5.1: Three-layered neural network

5.2 Training Data

We assume that the number of categories to be separated is l and that of training data in each category is m_τ, $\tau = 1, 2, \cdots, l$. We introduce the set

$$J_k = \{\nu \mid \sum_{j=0}^{k-1} m_j < \nu \leq \sum_{j=0}^{k} m_j\}, \qquad m_0 = 0.$$

The training data are denoted by x^ν, $\nu = 1, 2, \ldots, m$, where $m = \sum_{\tau=1}^{l} m_\tau$. We define the function $q(\nu)$ by

$$q(\nu) = k, \quad \nu \in J_k, \quad k = 1, 2, \cdots, l.$$

In the case that the number of training data in each category is the same, denoting it by s, J_k can be simply written as

$$J_k = \{\nu \mid \sum_{j=0}^{k-1} m_j < \nu \leq \sum_{j=0}^{k} m_j\}$$

$$
\begin{aligned}
&= \{\nu \mid s(k-1) < \nu \leq sk\} \\
&= \{\, s(k-1)+1, \cdots, sk \,\} \\
&= \{\nu \mid k = [\frac{\nu - 1}{s}] + 1\}.
\end{aligned}
$$

In the paper [21], this case has been treated.

5.3 Domains of Recognition

We impose on $\varphi_i(x)$ defined in (5.1) the following supervised conditions

$$\varphi_i(x^\nu) \geq 1-\varepsilon, \qquad i = q(\nu), \tag{5.2}$$

$$\varphi_i(x^\nu) \leq \varepsilon, \qquad i \neq q(\nu). \tag{5.3}$$

Since the function $s = g(t)$ is monotonically increasing and has the inverse function $g^{-1}(s) = \ln(s/(1-s))$, we can rewrite (5.2) and (5.3) as follows:

$$W_i \cdot \psi(x^\nu) - \theta_i \geq \ln \frac{1-\varepsilon}{\varepsilon}, \qquad i = q(\nu), \tag{5.4}$$

$$W_i \cdot \psi(x^\nu) - \theta_i \leq -\ln \frac{1-\varepsilon}{\varepsilon}, \qquad i \neq q(\nu). \tag{5.5}$$

We define a domain which is represented by using the same symbol $D_\rho(x^\nu)$ as in Section 3.3:

$$
\begin{aligned}
D_\rho(x^\nu) = \{x \in R^n \mid\ & W_i \cdot (\psi(x) - \psi(x^\nu)) \leq \rho \mid W_i \cdot \psi(x^\nu) - \theta_i \mid, \quad i \neq q(\nu), \\
& W_i \cdot (\psi(x) - \psi(x^\nu)) \geq -\rho \mid W_i \cdot \psi(x^\nu) - \theta_i \mid, \quad i = q(\nu)\},
\end{aligned}
$$

where $0 < \rho < 1$.

Although the domain $D_\rho(x^\nu)$ has a complicated shape, we have the following theorem.

Theorem 5.1. For any x in $D_\rho(x^\nu)$, we have

$$\varphi_i(x) \geq 1-\varepsilon^{1-\rho}, \qquad i = q(\nu), \tag{5.6}$$

$$\varphi_i(x) \leq \varepsilon^{1-\rho}, \qquad i \neq q(\nu). \tag{5.7}$$

Proof. Since the proof is essentially the same as the proof of Theorem 3.1, we describe only an outline of the proof. By the definition of $D_\rho(x^\nu)$ and in the same way as in the proof of Theorem 3.1, we have for $i = q(\nu)$,

$$\begin{aligned} W_i \cdot \psi(x) - \theta_i &= W_i \cdot (\psi(x) - \psi(x^\nu)) + W_i \cdot \psi(x^\nu) - \theta_i \\ &\geq (1-\rho)(W_i \cdot \psi(x^\nu) - \theta_i) \\ &\geq (1-\rho) \ln \frac{1-\varepsilon}{\varepsilon}. \end{aligned}$$

This inequality and the monotonicity of $g(t)$ lead us to

$$g(W_i \cdot \psi(x) - \theta_i) \geq g((1-\rho) \ln \frac{1-\varepsilon}{\varepsilon}). \tag{5.8}$$

Applying the inequality

$$(1-\rho) \ln \frac{1-\varepsilon}{\varepsilon} \geq \ln \frac{1-\varepsilon^{1-\rho}}{\varepsilon^{1-\rho}}$$

to (5.8), we have

$$\begin{aligned} g(W_i \cdot \psi(x) - \theta_i) &\geq g(\ln \frac{1-\varepsilon^{1-\rho}}{\varepsilon^{1-\rho}}) \\ &= 1 - \varepsilon^{1-\rho} \end{aligned}$$

which implies (5.6). Similarly, we can prove (5.7).

This theorem means that any x belonging to $D_\rho(x^\nu)$ can be recognized as x^ν. We call $D_\rho(x^\nu)$ a domain of recognition.

Furthermore, we can prove the following result.

Theorem 5.2. We define l unions of $D_\rho(x^\nu)$ by

$$S_k = \cup_{\nu \in J_k} D_\rho(x^\nu), \quad k = 1, 2, \cdots, l.$$

Then S_k are mutually disjoint. The S_k represent categories of classification as shown in Fig. 5.2.

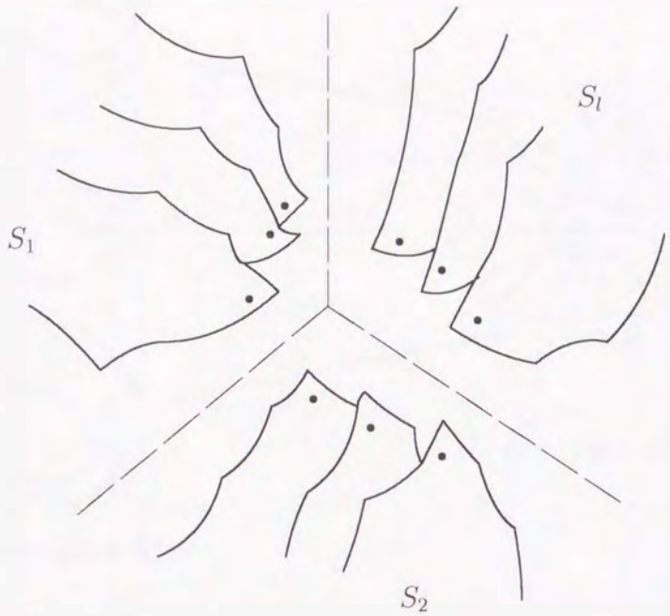


Fig. 5.2: The unions $S_k = \cup_{\nu \in J_k} D_\rho(x^\nu)$

Proof. The proof is by proof of contradiction. Suppose that $S_k \cap S_{k'} \neq \phi$ for $k \neq k'$, where ϕ denotes an empty set. Then there exists x^* belonging to $D_\rho(x^\nu)$ for $\nu \in J_k$ and $D_\rho(x^\nu)$ for $\nu \in J_{k'}$. Since J_k and $J_{k'}$ are different, there exists i such that $\varphi_i(x^*) \geq 1 - \varepsilon^{1-\rho}$ and $\varphi_i(x^*) \leq \varepsilon^{1-\rho}$. This leads us to a contradiction.

5.4 Learning Method

It is desirable for $D_\rho(x^\nu)$ to be large. Since the domain $D_\rho(x^\nu)$ takes a complicated form, it is difficult to make large $D_\rho(x^\nu)$ directly.

We consider a mapping of $D_\rho(x^\nu)$ into the hidden space R^h, which is given by

$$E_\rho(\psi(x^\nu)) = \{u \in R^h \mid W_i \cdot (u - \psi(x^\nu)) \leq \rho \mid W_i \cdot \psi(x^\nu) - \theta_i \mid, \quad i \neq q(\nu),$$
$$W_i \cdot (u - \psi(x^\nu)) \geq -\rho \mid W_i \cdot \psi(x^\nu) - \theta_i \mid, \quad i = q(\nu)\}. \tag{5.9}$$

We see from (5.9) that the boundaries of $E_\rho(\psi(x^\nu))$ consist of hyperplanes. From the

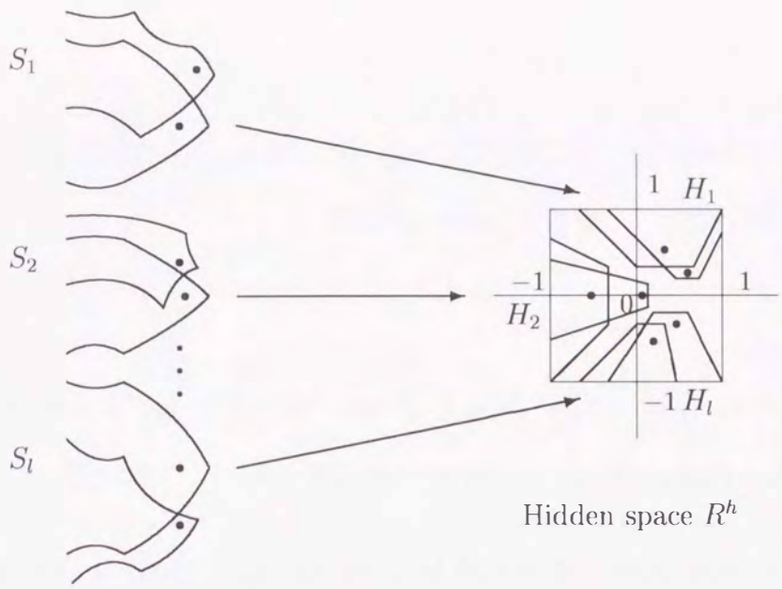


Fig. 5.3: The unions $S_k = \cup_{\nu \in J_k} D_\rho(x^\nu)$ and the unions $H_k = \cup_{\nu \in J_k} E_\rho(\psi(x^\nu))$

property of the mapping $u = \psi(x)$, the range $E_\rho(\psi(x^\nu))$ should be considered in the h-dimensional cube $[-1, 1]^h$. For latter use, we define l unions of $E_\rho(\psi(x^\nu))$ by

$$H_k = \cup_{\nu \in J_k} E_\rho(\psi(x^\nu)), \quad k = 1, 2, \cdots, l.$$

We illustrate the relation between the unions S_k and the unions H_k in Fig. 5.3.

Concerning the union H_k, we have the following result.

Corollary 5.3. At most one H_k includes zero in the hidden space R^h.

Proof. The proof is obvious from Theorem 5.2.

Corollary 5.3 will be used to enlarge $E_\rho(\psi(x^\nu))$ in the cube $[-1, 1]^h$.

Although $D_\rho(x^\nu)$ has a complicated form, $E_\rho(\psi(x^\nu))$ is a region whose boundaries consist of hyperplanes. To enlarge $D_\rho(x^\nu)$, we first make large the region $E_\rho(\psi(x^\nu))$. We decompose $E_\rho(\psi(x^\nu))$ into a cone $Cone(\psi(x^\nu))$ and a stripe $Str(\psi(x^\nu))$ as

$$E_\rho(\psi(x^\nu)) = Cone(\psi(x^\nu)) \cup Str(\psi(x^\nu)),$$

where

$$Cone(\psi(x^\nu)) = \Big\{ u \in R^h \mid W_i \cdot (u - \psi(x^\nu)) \leq 0, \quad i \neq q(\nu) \ , \\ W_i \cdot (u - \psi(x^\nu)) \geq 0, \quad i = q(\nu) \Big\} \tag{5.10}$$

and

$$Str(\psi(x^\nu)) = \Big\{ u \in R^h \mid 0 < W_i \cdot (u - \psi(x^\nu)) \leq \rho |W_i \cdot \psi(x^\nu) - \theta_i|, \quad i \neq q(\nu), \\ 0 > W_i \cdot (u - \psi(x^\nu)) \geq -\rho |W_i \cdot \psi(x^\nu) - \theta_i|, \quad i = q(\nu) \Big\}$$

As was done in [14], we make large the width of $Str(\psi(x^\nu))$, which can be expressed as

$$\frac{\rho |W_i \cdot \psi(x^\nu) - \theta_i|}{\|W_i\|}. \tag{5.11}$$

However, the quantity (5.11) depends on the index ν and so, $\sum_{\nu=1}^{m} (W_i \cdot \psi(x^\nu) - \theta_i)^2 / \|W_i\|^2$ is maximized, that is,

$$\frac{\|W_i\|^2}{\sum_{\nu=1}^{m} (W_i \cdot \psi(x^\nu) - \theta_i)^2} \tag{5.12}$$

is minimized.

Next, we minimize the angle γ at which the hyperplanes $W_i \cdot (u - \psi(x^\nu)) = 0$ and $W_j \cdot (u - \psi(x^\nu)) = 0$ cross. To do so, it suffices to minimize

$$\cos \gamma = \frac{W_i \cdot W_j}{\|W_i\| \, \|W_j\|}.$$

By the method in [15], we can expand $E_\rho(\psi(x^\nu))$ under the restrictions (5.4) and (5.5). Furthermore, by Corollary 5.3, $\|\psi(x^\nu)\|^2$ are desirable to be close to zero. In addition to the enlargement of $E_\rho(\psi(x^\nu))$ in the cube $[-1, 1]^h$, we minimize $\|V_j\|$ because the smaller the slope of affine transforms between the input and hidden layers is, the larger $D_\rho(x^\nu)$ becomes. Summarizing the above discussions, we minimize the cost function:

$$\sum_{i=1}^{l} \frac{\|W_i\|^2}{\sum_{\nu=1}^{m}(W_i\cdot\psi(x^\nu)-\theta_i)^2} + C_1 \sum_{i\neq j} \frac{W_i\cdot W_j}{\|W_i\|\,\|W_j\|}$$
$$+C_2 \sum_{\nu=1}^{m} \|\psi(x^\nu)\|^2 + C_3 \sum_{j=1}^{h} \|V_j\|^2$$
$$+C_4 \left[\sum_{i\neq q(\nu)} \left(\ln \frac{1-\varepsilon}{\varepsilon} + W_i\cdot\psi(x^\nu) - \theta_i \right)_+^2 \right.$$
$$\left. + \sum_{i=q(\nu)} \left(\ln \frac{1-\varepsilon}{\varepsilon} - W_i\cdot\psi(x^\nu) + \theta_i \right)_+^2 \right], \tag{5.13}$$

where C_i denote penalty constants, the function z_+^2 is defined in Section 3.4. By bounding the last penalty term, the conditions (5.4) and (5.5) become to be satisfied. Our learning algorithm is given as a minimizing process of this cost function.

In actual computation, however, we minimize the following functional in place of (5.13) to avoid numerical instability:

$$\sum_{i=1}^{l} \frac{1}{\sum_{\nu=1}^{m}(U_i\cdot\psi(x^\nu)-\xi_i)^2} + C_1 \sum_{i\neq j} U_i\cdot U_j$$
$$+C_2 \sum_{\nu=1}^{m}\sum_{j=1}^{h} (V_j\cdot x^\nu - \eta_j)^2 + C_3 \sum_{j=1}^{h} \|V_j\|^2$$
$$+C_4 \left[\sum_{i\neq q(\nu)} (\alpha + \beta_i(U_i\cdot\psi(x^\nu)-\xi_i))_+^2 \right.$$
$$\left. + \sum_{i=q(\nu)} (\alpha - \beta_i(U_i\cdot\psi(x^\nu)-\xi_i))_+^2 \right]$$
$$+C_5 \sum_{i=1}^{l} \left(\|U_i\|^2 - 1 \right)^2, \tag{5.14}$$

where we have put

$$\begin{aligned} U_i &= \frac{W_i}{\|W_i\|}, \\ \xi_i &= \frac{\theta_i}{\|W_i\|}, \\ \beta_i &= \|W_i\| \end{aligned}$$

and

$$\alpha = \ln \frac{1-\varepsilon}{\varepsilon}.$$

We use the steepest descent method as a minimization technique. This minimization process gives our learning algorithm.

Chapter 6

Land Cover Classification by LDR Method

Land cover classification is a typical problem in remote sensing. In this chapter, we apply the learning method proposed in Chapter 5 to make a land cover classification map. In simulations, we compare the results by our method with those by the maximum likelihood method.

6.1 Input Data and Training Data

We apply the learning method based on domains of recognition proposed in Chapter 5 for land cover classification by using TM and SAR data. We prepare a vector x consisting of 7 or 8 band values which are from visible to infrared reflectances observed by TM, and 1 band value of microwave scattering observed by AMI. The vector x takes the form $x = (x_1, x_2, \cdots, x_7)$ or $x = (x_1, x_2, \cdots, x_8)$. Each of $x_1, x_2, \cdots, x_7$ has 8 bits expression. However, since AMI data I have 16 bits expression, they were converted into the 8 bits data x_8 by the following equations:

$$\begin{aligned} \sigma_0 &= 20\times \log_{10}(I) - 68.5 \quad (dB), \\ x_8 &= 5(\sigma_0 + 30) \end{aligned}$$

in which σ_0 is called a backscattering coefficient.

To get a high ratio of separation, we transform the vector x into other vector by using the principle component analysis.

The procedure of the principle component analysis is as follows:
The principle components z_k are expressed as

$$z_k = a_{1k}x_1 + a_{2k}x_2 + \cdots + a_{8k}x_8. \tag{6.1}$$

The coefficients a_{ik} are determined so that

1. $\sum_{i=1}^{8} a_{ik}^2 = 1$ is satisfied,

2. The variance $\sigma_{z_k}^2$ is maximized,

3. z_k are independent of each other.

6.2 Land Cover Classification

6.2.1 Simulation I

Applying the learning method in Chapter 5, we carried out a land cover classification of Fukuoka area with the range about 46 by 38 kilometers, in Japan. Input data of the neural network consist of orthogonal components which are computed using the principle component analysis from 7 band data of Landsat 5 TM whose image is shown in Fig. 6.1. The principle components are illustrated in Fig. 6.2. We here use 4 principle components to construct an input vector. Categories to be classified are paddy, grassland, forest, urban and residential area, bare soil, and sea and river as shown in Fig. 6.3. However, since the land cover in Fukuoka area is complicated, we subdivided forest into 3 subcategories, urban and residential area into 2 subcategories, and bare soil into 2 subcategories. Therefore, the number of categories is 10. Training data are chosen from the data whose categories are known in advance.

The structure of the neural network is as follows: the number n of input units is 4, the number h of hidden units is 5, and the number l of output units is 10 which corresponds to the number of categories. The number of training data in each category is $m_1 = m_2 = \cdots = m_{10} = 10$. The penalty constants in the cost function (5.14) were selected as $C_1 = 5$, $C_2 = 15$, $C_3 = 1$ and $C_4 = 50$. As initial values of the weights W_i and V_j, random numbers were chosen. Initial values of thresholds θ_i and η_j, and ε were selected as $\theta_i = 5$, $\eta_j = 0$ and $\varepsilon = 10^{-2}$. We used the steepest descent method as a minimizing process of the cost function (5.14). By the learning of neural network, we could get 10 unions each of which consists of several domains of recognition.

The land cover classification map shown in Fig. 6.3 was constructed by estimating which union each pixel of the image data in Fukuoka area is contained in. We succeeded to classify 89% pixels in the image data into 10 categories. As shown in Table 6.1, we could get the ratios of land cover which are extremely similar to those computed based on the land use map (Fig. 6.4) of Digital National Land Information.

Fig. 6.1: Landsat 5 TM false color composite image in Fukuoka area, Japan, acquired on April 24, 1997. The image displays band 5 as red, band 4 as green and band 3 as blue. Bird's eye view.

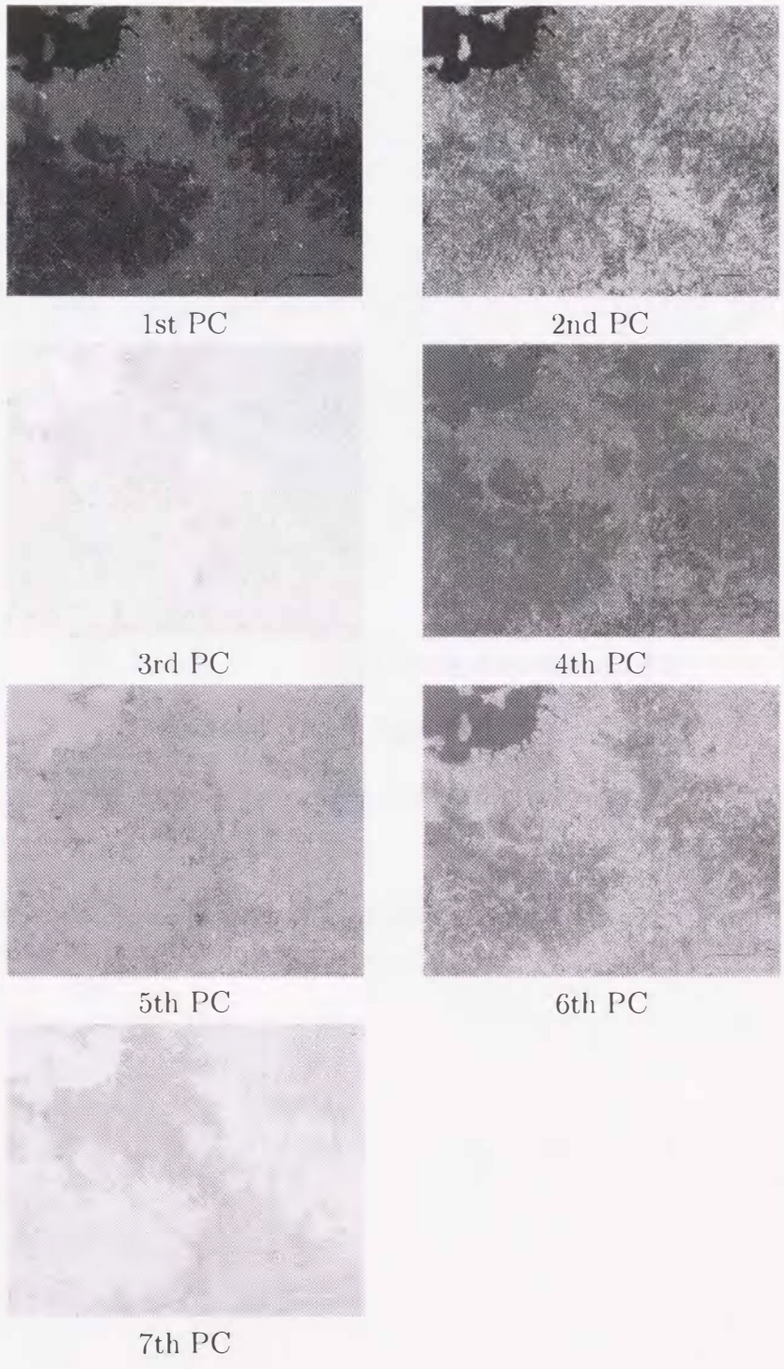


Fig. 6.2: Images of input data consisting of principle components (PC) for the image shown in Fig. 6.1

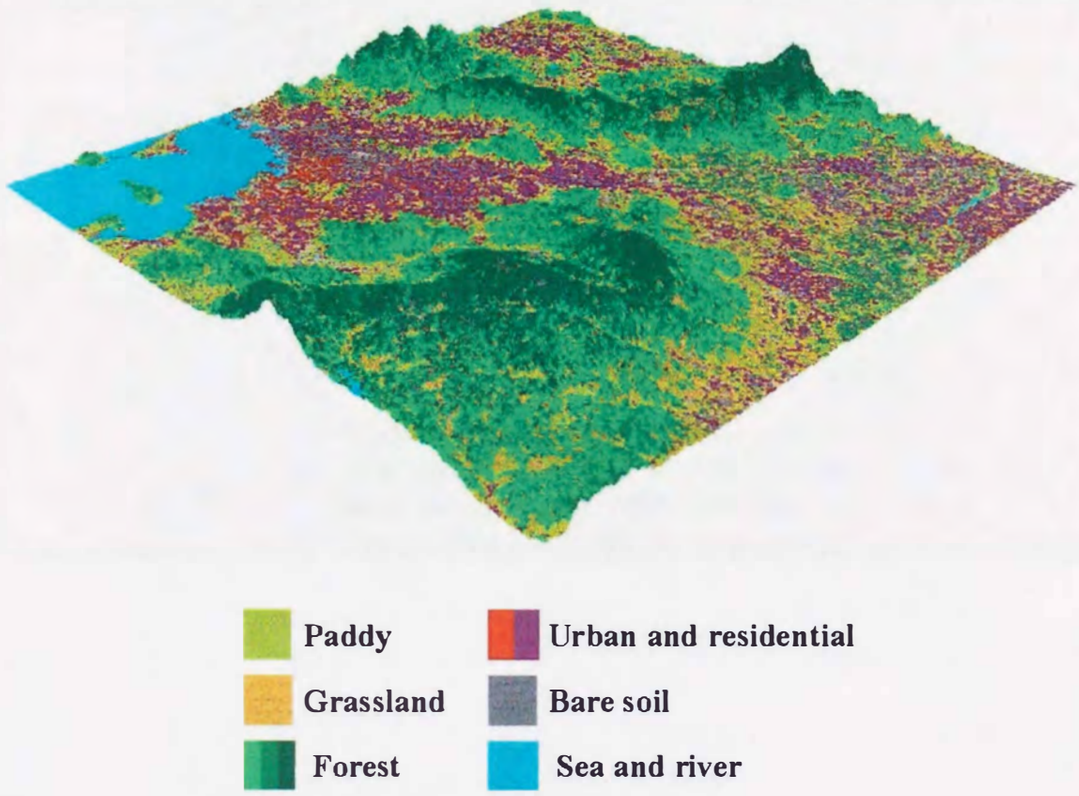


Fig. 6.3: Land cover classification map produced by LDR method for TM image in Fig. 6.1, displaying by bird's eye view.

Table 6.1: Ratios of land cover obtained by LDR method and land use map in Fukuoka area (without SAR data)

	Category	LDR method	Land use[1),2)]
1	Paddy	20.2	20.7
2	Grassland	6.8	2.8
3	Forest	45.4	45.5
4	Urban and residential area	16.5	18.8
5	Bare soil	3.6	4.6
6	Sea and river	7.5	7.6
	Total	100.0[3)]	100.0

1) Digital National Land Information.
2) The values for 1,2,4,5 and 6 indicate total values of some subcategories.
3) Unclassified pixels are not included.

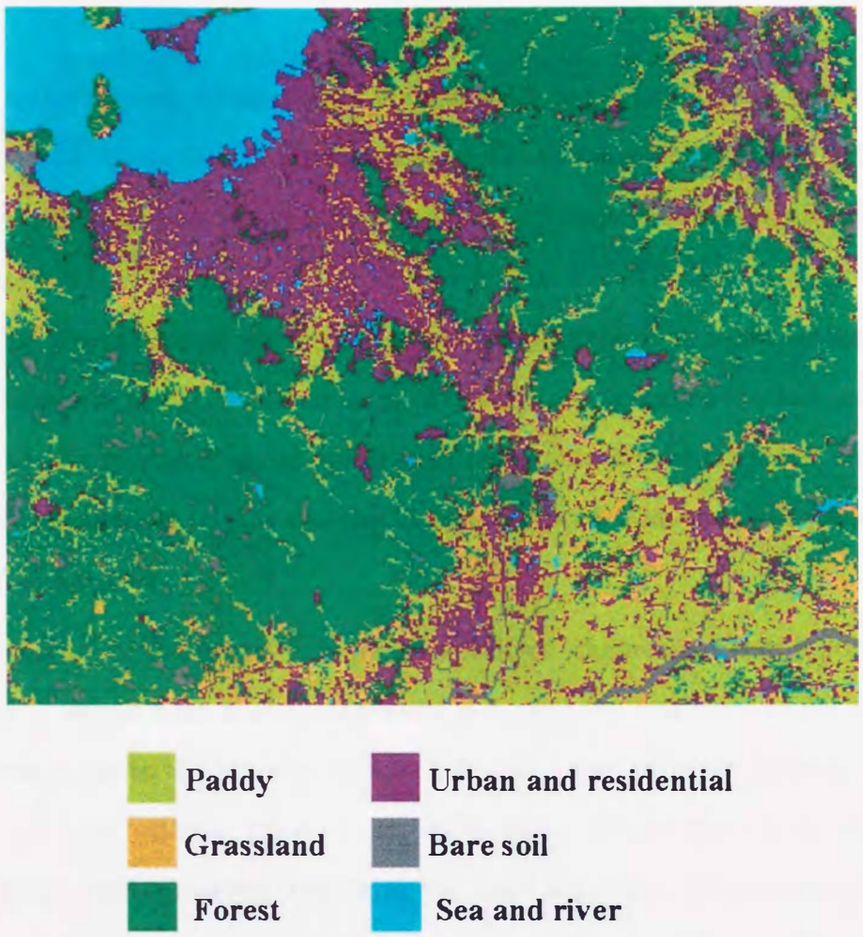


Fig. 6.4: Land use map produced by Digital National Land Information

6.2.2 Simulation II

The next objective area of land cover classification is Chikushi plain in Kyushu Island, Japan (Fig. 6.5). Input data of the neural network consist of orthogonal components as in Simulation I, which are sought by using the principle component analysis from 7 band data of Landsat 5 TM. The TM data are imaged in Fig. 6.5. The principle components are illustrated in Fig. 6.6, all of which are used as components of an input vector. Categories to be classified are paddy, field and orchard, forest, urban and residential area, bare soil, and sea and river. Moreover, we subdivided forest into 2 subcategories. Therefore, the number of categories is 7. The training data are chosen from the data whose categories are known in advance.

The structure of the neural network is as follows: the number n of input units of the network is 7, the number h of hidden units is 5, and the number l of output units is 7 which corresponds to the number of categories. For each of these categories, we choose $m_1 = 13$, $m_2 = 12$, $m_3 = 9$, $m_4 = 16$, $m_5 = 16$, $m_6 = 16$, and $m_7 = 11$ as the number of training data in each category. Therefore, the total number m of training data is 93. The penalty constants in the cost function (5.14) were selected as $C_1 = 5$, $C_2 = 15$, $C_3 = 1$, $C_4 = 50$ and $C_5 = 0.5$. As initial values of the weights W_i and V_j, random numbers were chosen. Initial values of thresholds θ_i and η_j, and ε were selected as $\theta_i = 5$, $\eta_j = 0$ and $\varepsilon = 10^{-2}$. We used the steepest descent method as a minimizing process of (5.14). By the learning of neural network, we could obtain 7 unions each of which consists of several domains of recognition.

The land cover classification map shown in Fig. 6.7 was constructed by estimating which union each pixel of the image data in Chikushi plain is contained in. We succeeded to classify 90.5% pixels in the image data into 7 categories. As shown in Table 6.2, we could get the ratios of land cover which are similar to those computed based on the land

use map of Digital National Land Information (Fig. 6.8).

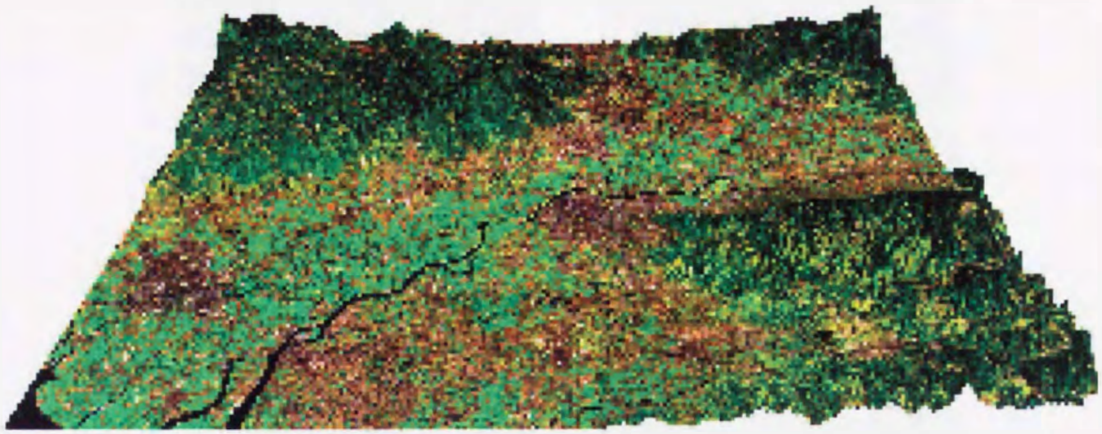


Fig. 6.5: Landsat 5 TM false color composite image in Chikushi plain, Japan, acquired on April 24, 1997. The image displays band 5 as red, band 4 as green and band 3 as blue. Bird's eye view.

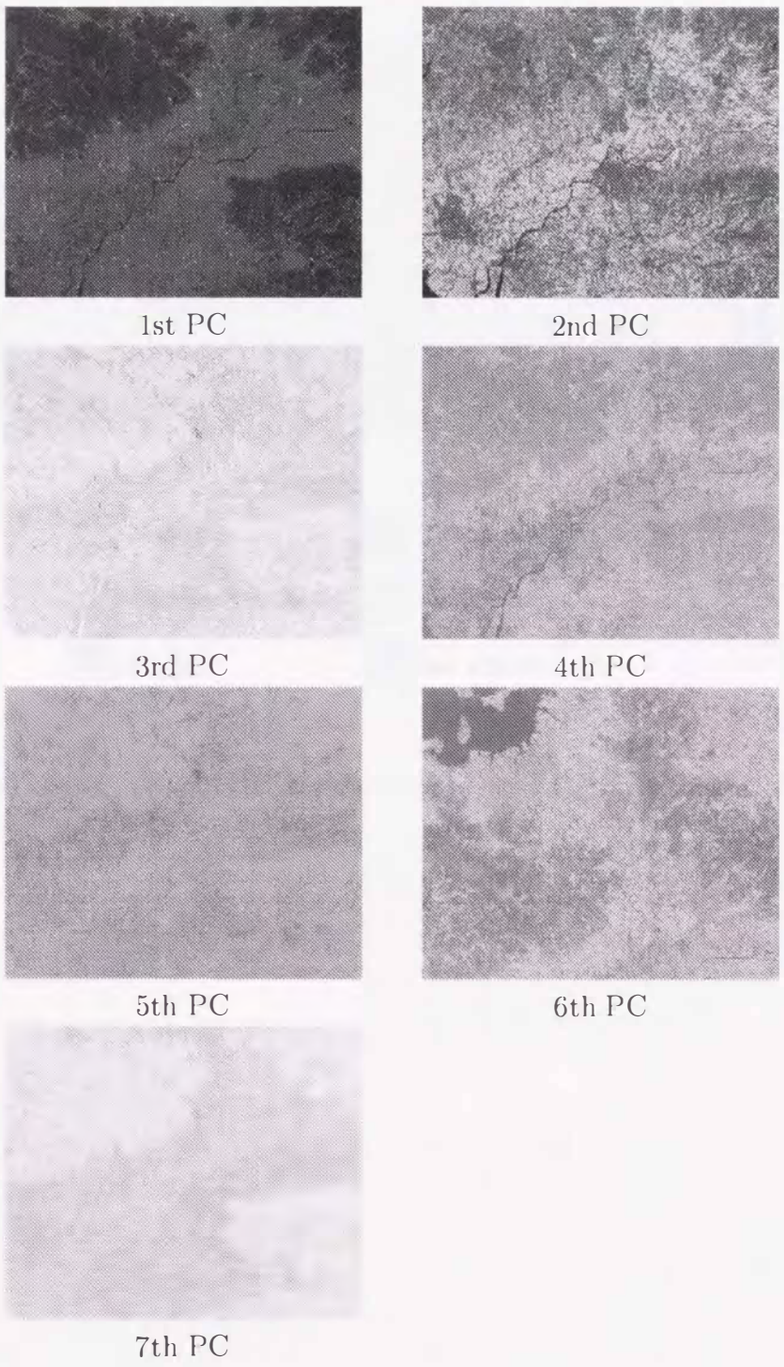


Fig. 6.6: Images of input data consisting of the principle components (PC) for the image shown in Fig. 6.5

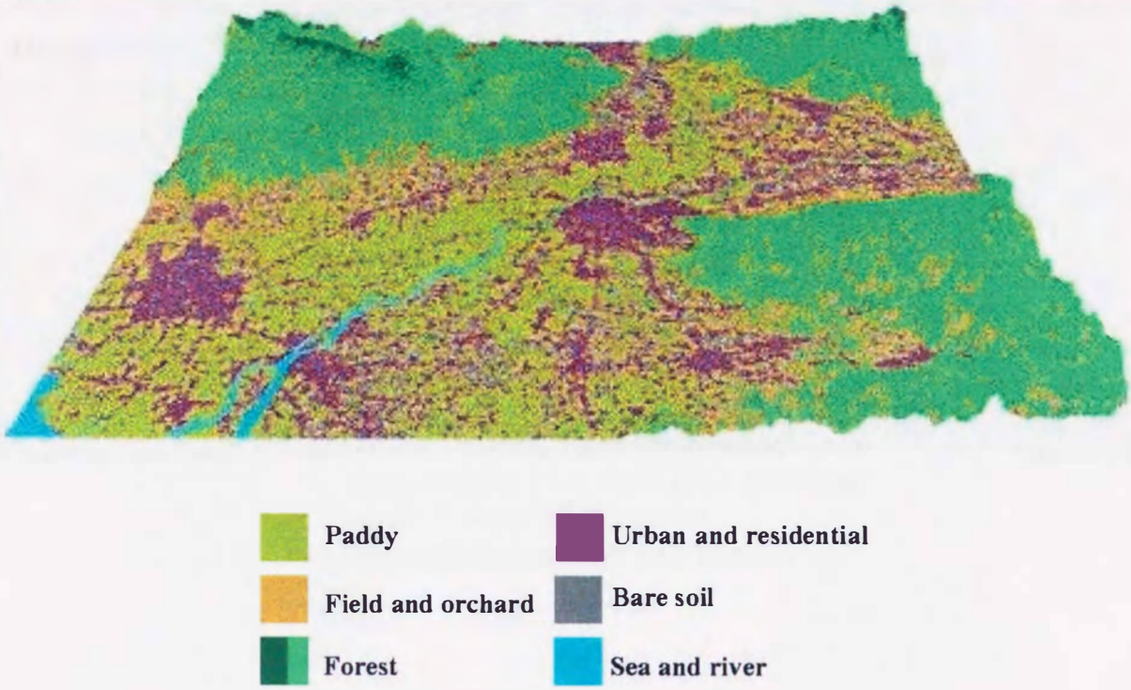


Fig. 6.7: Land cover classification map produced by LDR method for TM image in Fig. 6.5, displaying by bird's eye view.

Table 6.2: Ratios of land cover obtained by LDR method and land use map in Chikushi plain (without SAR data)

	Category	LDR method	Land use
1	Paddy	39.3	34.5
2	Field and orchard	10.5	8.4
3	Forest	30.0	37.7
4	Urban and residential area	16.4	14.1
5	Bare soil	3.0	4.3
6	Sea and river	0.8	1.0
	Total	100.0 [3)]	100.0

1) Digital National Land Information.
2) The values for 1,3,4,5 and 6 indicate total values of some subcategories.
3) Unclassified pixels are not included.

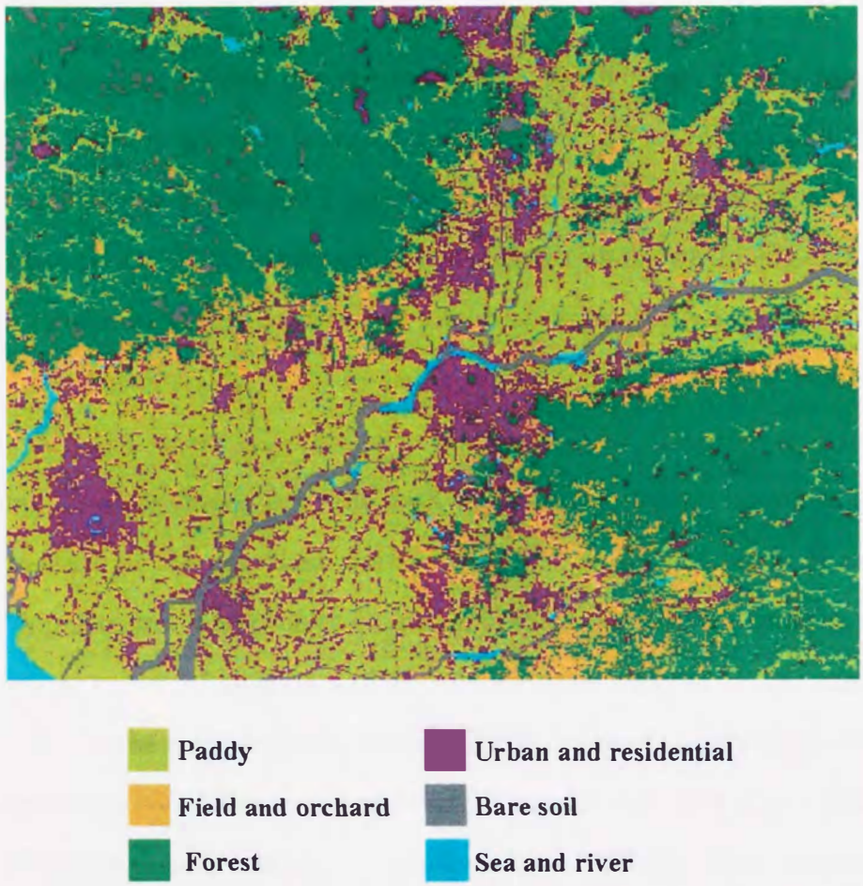


Fig. 6.8: Land use map produced by Digital National Land Information

6.2.3 Simulation III

In this simulation, we carried out the land cover classification for the same area as in Simulation II, but now we used 8 band data consisting of 7 band data of TM and 1 band data of AMI. Images constructed from AMI data in the area where the altitude is higher than $70\,m$ are prone to be influenced by mountains as shown in Fig. 6.9. Therefore, we eliminated the mountain area in our analysis. From the 8 band data, we compute orthogonal components (Fig. 6.10) by using the principle component analysis to get an 8-dimensional input vector. Categories to be classified are the same as in Simulation II. The training data are chosen from the data whose categories are known in advance.

The structure of the neural network is the same as in Simulation II except for the number of input nodes. The penalty constants in (5.14) are also the same as in Simulation II. The initial values of weights and thresholds were selected in the same way as in Simulation II. In the same way as in Simulation II, we used the steepest descent method to minimize (5.13). As a result, we could get 7 unions each of which consists of several domains of recognition. Classification was done by estimating which union each pixel of the image data is contained in. We succeeded to classify 92.3% pixels in the image data into 7 categories (Fig. 6.11). As shown in Table 6.3, we could get the similar ratios of land cover to those computed based on Digital National Land Information in Fig. 6.12. Comparing the classification map obtained by adding AMI data with that in Simulation II, unclassified pixels of the former are less than those of the latter.

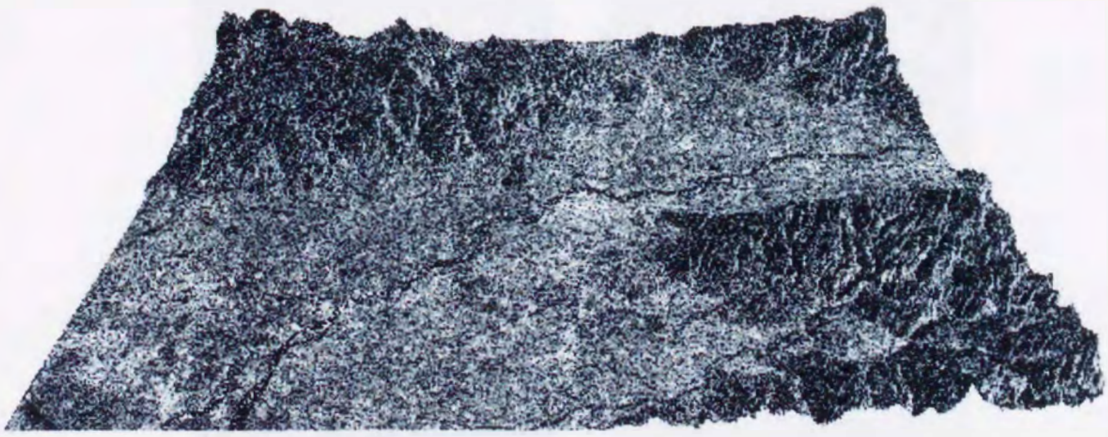


Fig. 6.9: ERS-2 AMI image in Chikushi plain, Japan, acquired on January 17, 1997. The image displays back scatter with a gray scale. Bird's eye view.

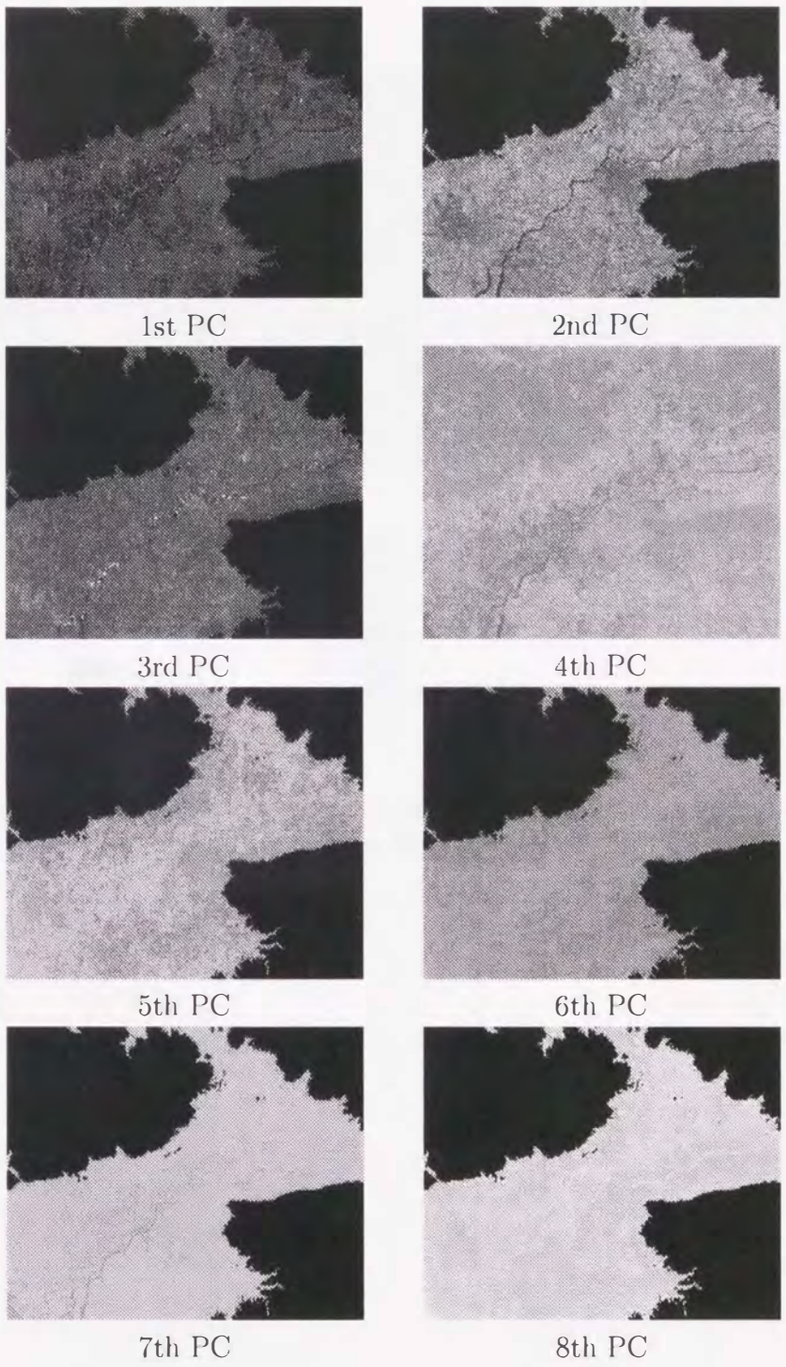


Fig. 6.10: Images of input data consisting of principle components (PC) for the images shown in Fig. 6.5 and Fig. 6.9

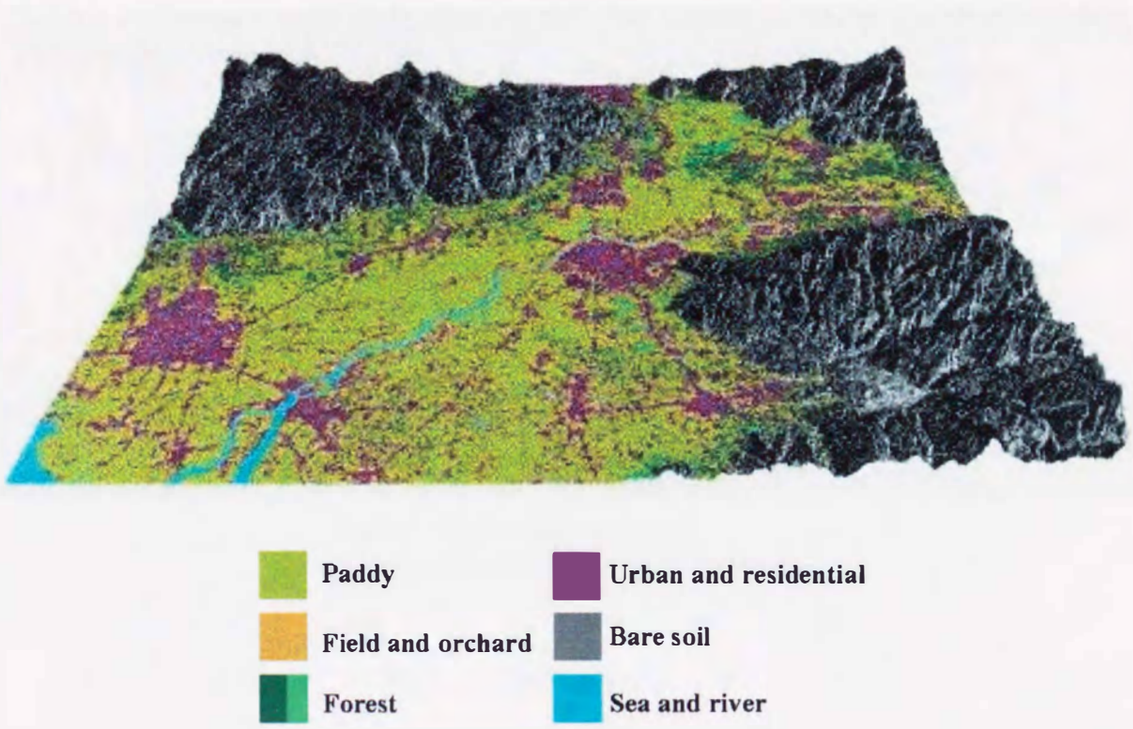


Fig. 6.11: Land cover classification map produced by LDR method for TM and AMI images in Fig. 6.5 and Fig. 6.9, displaying by bird's eye view

Table 6.3: Ratios of land cover obtained by LDR method and land use map in Chikushi plain (with SAR data)

	Category	LDR method	Land use
1	Paddy	56.2	57.5
2	Field and orchard	4.9	6.3
3	Forest	15.3	5.8
4	Urban and residential area	19.5	23.7
5	Bare soil	2.6	5.3
6	Sea and river	1.5	1.6
	Total	100.0 [3)]	100.0

1) Digital National Land Information.
2) The values for 1,3,4,5 and 6 indicate total values of some subcategories.
3) Unclassified pixels are not included.

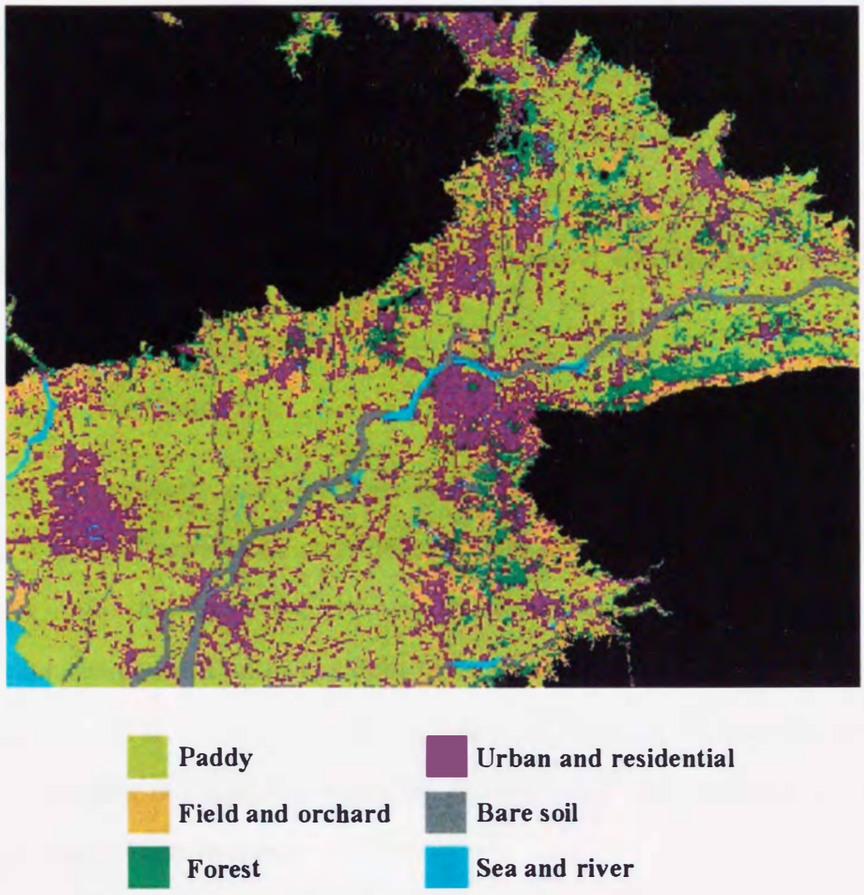


Fig. 6.12: Land use map produced by Digital National Land Information

6.3 Land Cover Classification by Maximum Likelihood (ML) Method

To compare the classification ability of our method with that of ML method, we tried to classify the same object areas as in Simulations I, II and III by using ML method. The classification results by ML method were also compared with the land use map of Digital National Land Information.

6.3.1 ML Method

ML method is frequently used for classifying the remote sensing data. This statistical method is a kind of supervised classification, and unknown data are classified based on the training data that have already been known what category they belong to. It is assumed that the population of the data has a multi-dimensional normal distribution.

Let $x_{jk}^{\nu}\,;\, j = 1, 2, \cdots, n,\, k = 1, 2, \cdots, l,\, \nu = 1, 2, \cdots, m$ be training data, where m denotes the number of data, n the number of bands, and l the number of class. We first define an n-dimensional vector by

$$X_k^{\nu} = {}^t(\, x_{1k}^{\nu},\, x_{2k}^{\nu}, \cdots, x_{nk}^{\nu}),$$

and its average vector by

$$\overline{X_k} = \frac{1}{m}\sum_{\nu=1}^{m} X_k^{\nu}.$$

Using these vectors, we compute the covariance matrix

$$S_k = \frac{1}{m}\sum_{\nu=1}^{m}(X_k^{\nu} - \overline{X_k})\,{}^t(X_k^{\nu} - \overline{X_k})\,.$$

Then, the likelihood of k-th class for an unknown vector $X = {}^t(\, x_1,\, x_2, \cdots,\, x_n)$ except the training data can be expressed as

$$f_k(X) = \frac{1}{(2\pi)^{n/2}\,|\det S_k|^{1/2}} \exp\left\{-\frac{1}{2}{}^t(X - \overline{X_k})S_k^{-1}(X - \overline{X_k})\right\}.$$

We substitute a vector X to be discriminated into $f_k(X)$ and choose k such that $f_k(X)$ is maximal. The number k indicates a categorized class of X. Actually, we compute

$$g_k(X) = \log|\det S_k| + {}^t(X - \overline{X_k}){S_k}^{-1}(X - \overline{X_k})$$

and choose k such that $g_k(X)$ is minimal.

6.3.2 Classification by ML Method and Comparison with LDR Method

The classification by ML method was carried out using the same training data in Simulations I, II and III in Section 6.2. The classification maps are shown in Figs. 6.13, 6.14 and 6.15. Moreover, the classification results were listed in Tables 6.4, 6.5 and 6.6 together with those by LDR method and land use.

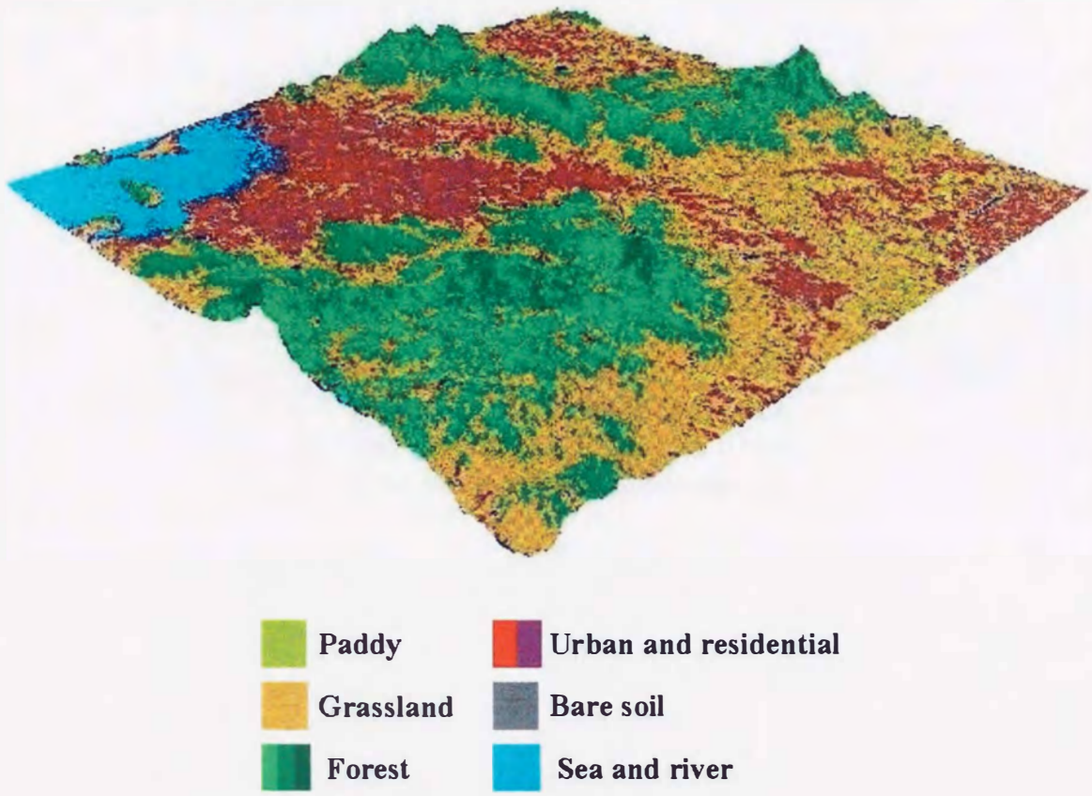


Fig. 6.13: Land cover classification map produced by ML method for TM image in Fig. 6.1, displaying by bird's eye view

Table 6.4: Ratios of land cover obtained by ML method, LDR method, and the land use map in Fukuoka area (without SAR data)

	Category	ML method	LDR method	Land use[1),2)]
1	Paddy	6.9	20.2	20.7
2	Grassland	31.2	6.8	2.8
3	Forest	32.0	45.4	45.5
4	Urban and residential area	21.3	16.5	18.8
5	Bare soil	3.2	3.6	4.6
6	Sea and river	5.4	7.5	7.6
	Total	100.0	100.0	100.0

1) Digital National Land Information.
2) The values for 1,2,4,5 and 6 indicate total values of some subcategories.

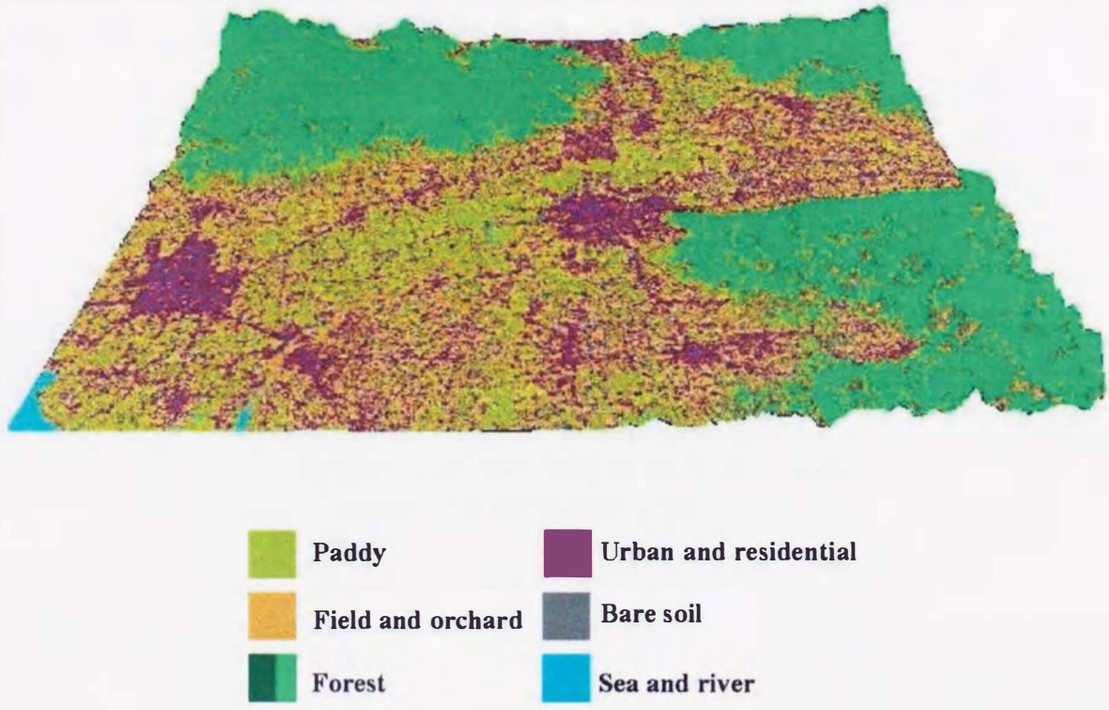


Fig. 6.14: Land cover classification map produced by ML method for TM image in Fig. 6.5, displaying by bird's eye view

Table 6.5: Ratios of land cover obtained by ML method, LDR method, and the land use map in Chikushi plain (without SAR data)

	Category	ML method	LDR method	Land use[1),2)]
1	Paddy	24.5	39.3	34.5
2	Field and orchard	22.3	10.5	8.4
3	Forest	30.5	30.0	37.7
4	Urban and residential area	16.3	16.4	14.1
5	Bare soil	6.2	3.0	4.3
6	Sea and river	0.2	0.8	1.0
	Total	100.0	100.0	100.0

1) Digital National Land Information.
2) The values for 1,3,4,5 and 6 indicate total values of some subcategories.

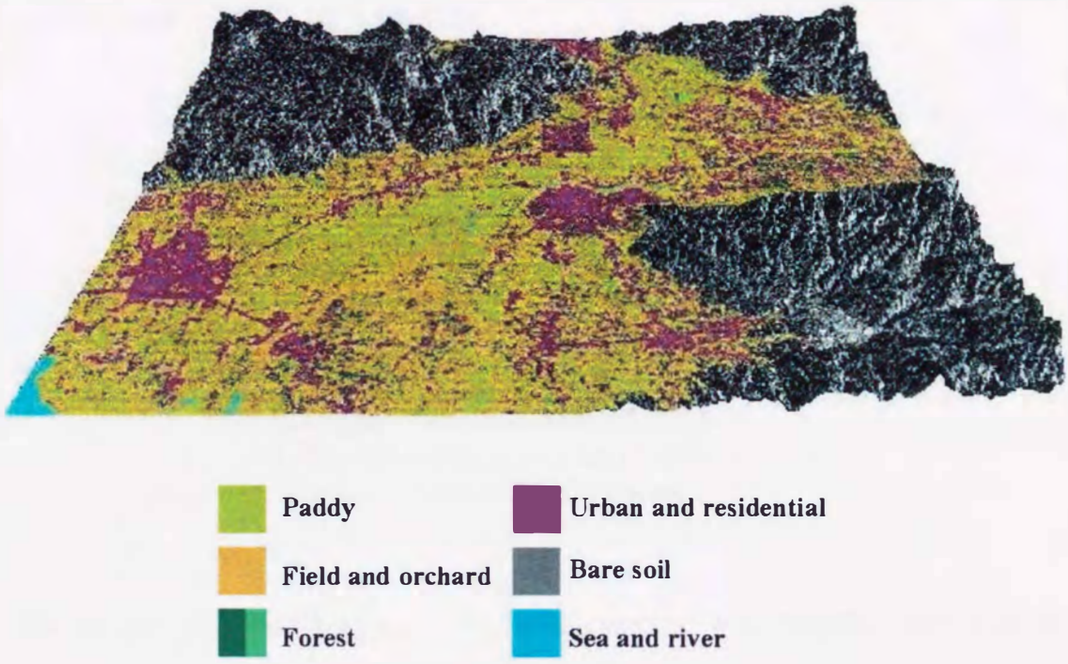


Fig. 6.15: Land cover classification map produced by ML method for adding AMI image in Fig. 6.9, displaying by bird's eye view

Table 6.6: Ratios of land cover obtained by ML Method, LDR method, and the land use map in Chikushi plain (with SAR data)

	Category	ML method	LDR method	Land use[1),2)]
1	Paddy	32.0	56.2	57.5
2	Field and orchard	38.7	4.9	6.3
3	Forest	8.4	15.3	5.8
4	Urban and residential area	19.9	19.5	23.7
5	Bare soil	0.5	2.6	5.3
6	Sea and river	0.5	1.5	1.6
	Total	100.0	100.0	100.0

1) Digital National Land Information.
2) The values for 1,3,4,5 and 6 indicate total values of some subcategories.

We compare the classification results by ML method with those by LDR method and the land use map of Digital National Land Information. In all cases and all categories, the ratios of land cover obtained by LDR method are very close to those of land use information. In the ML method, we had an overestimation of the field and orchard category and an underestimation in the paddy and forest categories, in comparison with the land use information.

6.4 Conclusion

We applied the LDR method to three land cover classification problems. As a result, we could obtain excellent classification results in all the simulations in comparison with the land use map. By adding SAR data, the accuracy of classification was improved.

We also applied ML method to obtain the ratios of land cover in the three cases. In comparison with the results by ML method, we see that LDR method is superior to ML method.

Chapter 7

General Conclusions

In this thesis, we have developed a learning theory of neural networks and applied it to the estimation of soil moisture and land cover classification.

As the first piece of research, we have proposed a learning method of three-layered neural networks, which is based on output errors minimization with successive addition of hidden units and cone-like domains of recognition. Moreover, we applied the learning technique to estimate soil moisture in the plain using the TM and SAR data and to make a soil moisture map. Simulation results show the superiority of our method in comparison with statistical methods. This learning method has a trade-off problem that if the number of hidden units increases, the output errors of the network becomes small while the domains of recognition become small.

As the second research subject, we have introduced the concept of domains of recognition, all of whose elements can be recognized as the training pattern in the domain. Furthermore, by using a simple shape of the region which is a mapping of the domain of recognition into the hidden space, we have developed a novel learning method that is an enlarging process of the region. Furthermore, we applied this learning method to get land cover classification for three object areas. We also compared the simulation results with those obtained by the maximum likelihood method.

The future works and remained problems are depicted as follows:

1. The cost functions to be minimized for learning neural networks contain many penalty constants to be adjusted. It is necessary to accomplish a method of tuning these free parameters.

2. We used in this thesis the steepest decent method to minimize the cost functions. This gradient method always has a problem of trapping to local minima. Global minimization techniques must be developed.

3. Our learning methods can be applied to other problems as well as estimation and classification problems in remote sensing.

Bibliography

[1] P.M. Atkinson and A.R.L. Tatnall, "Introduction : Neural networks in remote sensing," International Journal of Remote Sensing, 18, 4, pp.699-709, 1997.

[2] M.F. Augusteijn, L.E. Clemens and K.A. Shaw, "Performance evaluation of texture measures for ground cover identification in satellite images by means of a neural network classifier," IEEE Transactions on Geoscience and Remote Sensing, 33, 3, pp.616-626, 1995.

[3] U. Bhattacharya and S.K. Parui, "An improved backpropagation neural network for detection of road-like features in satellite imagery," International Journal of Remote Sensing, 18, 16, pp.3379-3394, 1997.

[4] J.B. Boisvert, Q.H.J. Gwyn, B. Brisco, D.J. Major and R.J. Brown, "Evaluation of soil moisture estimation techniques and microwave penetration depth for radar applications," Canadian Journal of Remote Sensing, 21, 2, pp.110-123, 1994.

[5] S. Cote and A.R.L. Tatnall, "The Hopfield neural network as a tool for feature tracking and recognition from satellite sensor image," International Journal of Remote Sensing, 18, 4, pp.871-885, 1997.

[6] J. Ediriwickrema and S. Khorram, "Hierarchical maximum likelihood classification for improved accuracies," IEEE Transactions on Geoscience and Remote Sensing, 35, 4, pp.810-816, 1997.

[7] O.C.R. Filho, E.S.N. Kouwen, A.A. Bdeh, K.O. Lahchi, T.J. Pultz and Y. Crevier, "Soil moisture in pasture field using ERS-1 SAR data: Preliminary results," Canadian Journal of Remote Sensing, 22, 1, pp.95-107, 1995.

[8] S. Gopal and C. Woodcock, "Remote sensing of forest change using artificial neural networks," IEEE Transactions on Geoscience and Remote Sensing, 34, 2, pp.398-404, 1996.

[9] Y. Ito and S. Omatu, "Category classification method using a self-organizing neural network," International Journal of Remote Sensing, 18, 4, pp.829-845, 1997.

[10] X. Jia and J.A. Richards, "Efficient maximum likelihood classification for imaging spectrometer data sets," IEEE Transactions on Geoscience and Remote Sensing, 32, 2, pp.274-281, 1994.

[11] M.H. Mohamed, T. Minamoto and K. Niijima, "Convergence rate of minimization learning for neural networks," Tenth European Conference on Machine Learning, Lecture Notes in Artificial Intelligence, 1398, Springer, pp.412-417, 1998.

[12] M.H. Mohamed, A. Ohkubo, T. Minamoto and K. Niijima, "Learning of three-layer neural networks based on domains of attraction," Proceedings of the Fifth International Conference on Neural Information Processing, pp.1591-1594, 1998.

[13] H. Murai and S. Omatu, "Remote sensing image analysis using a neural network and knowledge-based processing," International Journal of Remote Sensing, 18, 4, pp.811-828, 1997.

[14] K. Niijima, "Domains of attraction in autoassociative memory networks," New Generation Computing, 12, pp.395-407, 1994.

[15] K. Niijima, M. Yamada, M.H. Mohamed, T. Akanuma, T. Minamoto and A. Ohkubo, "Minimization learning of neural networks by adding hidden units," Research Reports on Information Science and Electrical Engineering of Kyushu University, 2, 2, pp.173-178, 1997.

[16] K. Niijima, "Learning of associative memory networks based upon cone-like domains of attraction," Neural Networks, 10, 9, pp.1649-1658, 1997.

[17] A. Ohkubo, K. Higashi, T. Okuda, K. Mori and Y. Yasuoka, "Drawing of land cover classification in city area of Fukuoka prefecture," Journal of the Remote Sensing Society of Japan, 11, 4, pp.77-81, 1991 (in Japanese).

[18] A. Ohkubo and Y. Yasuoka, "Annual change analysis of land cover in comparison with the land use map," Journal of the Remote Sensing Society of Japan, 16, 3, pp.65-76, 1996 (in Japanese).

[19] A. Ohkubo, M.H. Mohamed and K. Niijima, "A soil moisture map generated from satellite data by using domains of attraction in neural networks," Proceedings of the Fifth International Conference on Neural Information Processing, pp.356-359, 1998.

[20] A. Ohkubo, J. Takagi, N. Kuroyanagi, N. Hatae and M. Tamura, "The wide area estimation of soil moisture by artificial satellite data and the synchronized survey with satellite flying time," Journal of the Remote Sensing Society of Japan, 10, 1, pp.30-44, 1999 (in Japanese).

[21] A. Ohkubo and K. Niijima, "A new supervised learning method of neural networks and its application to the land cover classification," Proceedings of IEEE International Geoscience and Remote Sensing Symposium, pp.1369-1371, 1999.

[22] A. Ohkubo and K. Niijima, "New supervised learning of neural networks for satellite image classification," Proceedings of IEEE International Conference on Image Processing, pp.505-509, 1999.

[23] J.D. Paola and R.A. Schowengerdt, "A detailed comparison of backpropagation neural network and maximum likelihood classifiers for urban land use classification," IEEE Transactions on Geoscience and Remote Sensing, 33, 4, pp.981-996, 1995.

[24] J.T. Pulliainen, P.J. Mikkela, M.T. Hallikainen and J.P. Ikonen, "Seasonal dynamics of C-band backscatter of boreal forests with applications to biomass and soil moisture estimation," IEEE Transactions on Geoscience and Remote Sensing, 34, 3, pp.758-770, 1996.

[25] J.A. Richards and X. Jia, *"Remote sensing digital image analysis,"* Third, Revised and Enlarged Edition, Berlin, Springer, 1998.

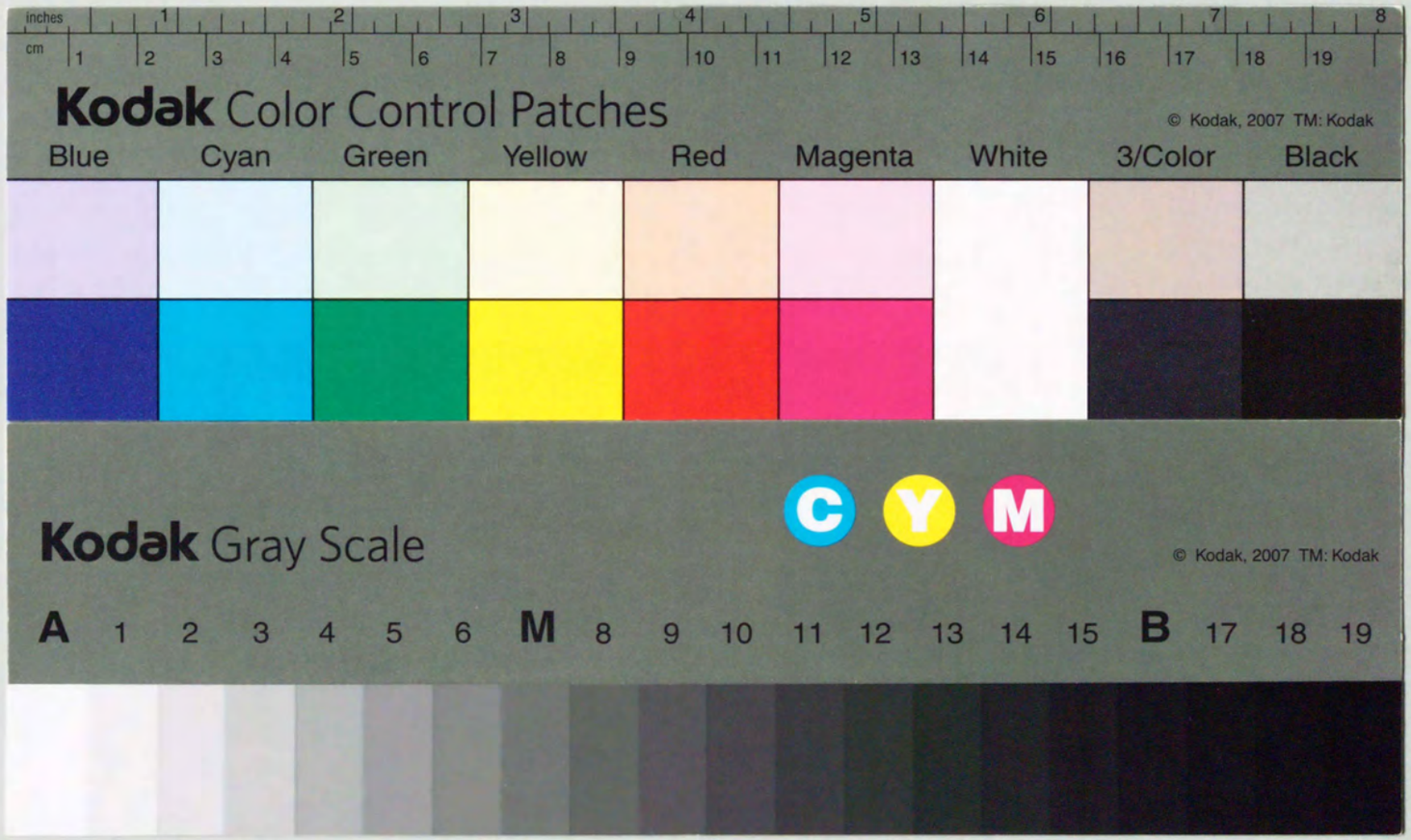
[26] D.E. Rumelhart and J.L. McClelland, *"Parallel distributed processing,"* Cambridge, Mass., MIT Press, 1986.

[27] S.B. Serpico and F. Roli, "Classification of multisensor remote sensing images by structured neural networks," IEEE Transactions on Geoscience and Remote Sensing, 33, 3, pp.562-578, 1995.

[28] D. Tsintikidis, J.L. Haferman, E.N. Anagnostou, W.F. Krajewski and T.F. Smith, "A neural network approach to estimating rainfall from spaceborne microwave data," IEEE Transactions on Geoscience and Remote Sensing, 35, 5, pp.1079-1093, 1997.

[29] S.R. Yhann and J.J. Simpson, "Application of neural networks to AVHRR cloud segmentation," IEEE Transactions on Geoscience and Remote Sensing, 33, 3, pp.590-604, 1995.

[30] T. Yoshida and S. Omatu, "Neural network approach to land cover mapping," IEEE Transactions on Geoscience and Remote Sensing, 32, 5, pp.1103-1109, 1994.


inches
cm
Kodak Color Control Patches
© Kodak, 2007 TM: Kodak
Blue
Cyan
Green
Yellow
Red
Magenta
White
3/Color
Black
Kodak Gray Scale
C
Y
M
© Kodak, 2007 TM: Kodak
A 1 2 3 4 5 6 M 8 9 10 11 12 13 14 15 B 17 18 19