Data Mining in Biomedical Imaging, Signaling, and Systems

Edited by
Sumeet Dua and Rajendra Acharya U.

Data Mining in Biomedical Imaging, Signaling, and Systems

Edited by
Sumeet Dua and Rajendra Acharya U.

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
AN AUERBACH BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20121102

International Standard Book Number-13: 978-1-4398-3939-3 (eBook - PDF)

Contents

Preface

Biosignals, medical images, and biosystems can effectively and efficiently characterize the health of the subject and can assist clinicians in making refined diagnoses for treatment. Because of the immense volume, heterogeneity, and complexity of data involved in these techniques, data mining has become an important amalgamation of indispensible and evolving tools for the accurate and fast implementation of biomedical applications. Novel data mining algorithms have aided in discovering and clarifying hidden information in medical data and have helped in accurately classifying pathological data from normal data. Data mining applied to medical data can help automated decision making, visualization, and extraction of hidden complex features from different patient groups and disease states. This book has 16 chapters that describe the applications of data mining in biosignals, medical imaging, and biosystems, and they explain how the applications aid in clinical diagnostic discovery. These chapters contain several examples of heterogeneous data modalities that are frequently encountered in biomedical and clinical applications, as well as detailing the applicability of fundamental data mining approaches and paradigms employed to address the computational challenges in analyzing this data for knowledge discovery.

The analysis of biosignals and medical images is significant in health care. Chapter 1 contains a description of the fundamental feature extraction methods employed in biomedical signaling and imaging. The chapter introduces feature extraction techniques for biomedical signaling, including techniques in frequency and statistical domains and information-theoretic methods. Such techniques are discussed in significant detail. A number of critical feature descriptors in biomedical signaling and imaging are also covered in this chapter.

Machine learning has been evaluated and employed in many biomedical applications for the diagnosis of diseases. The fundamentals of supervised learning and unsupervised learning are outlined in Chapter 2. Supervised and unsupervised learning methods, their performance evaluation measures, and the challenges that may have to be faced during their implementation for biosignals and medical images are also discussed in this chapter.

Depression is the world's fourth most serious health threat, and the incidence of this disease is expected to rise steadily with the aging of world population. Speech patterns of depressed speakers have often been characterized as dull, monotone, monoloud, lifeless, and metallic. These perceptual qualities have been associated with fluctuations involving fundamental frequency, formant structure, and power spectral distribution. Identifying whether individuals are affected with depression is often one of the most important judgments that physicians must make. Chapter 3 contains an analysis of depressed speech using acoustic properties, such as vocal jitter, formant frequencies, formant bandwidths, and power distribution. These parameters are extracted from dyadic wavelet transform (D_yWT) for the detection of normal and depressed speech. The results indicate the use of speech-processing tools for the detection and analysis of depressed subjects.

Biological systems are highly complex, and knowledge about them is fragmented. Thus, the creation and evaluation of clinical prediction rules are time consuming and expensive. However, with growing accessibility to electronically stored medical data and advances in data mining methods, the creation, evaluation, and utilization of predictive models can be supported by automated processes, which can induce models and patterns for large collections of patient medical records and medical data sets used in clinical trials. Chapter 4 contains an explanation of the concept of typicality from a broad perspective of cognitive psychology and of the intracategory and intercategory typicality measures; it presents a fuzzy set representation of typicality and a discussion on the applications of typicality measures used in visual classifications of objects.

Arrhythmia and ischemia are two of the most prominent life-threatening heart abnormalities. The automated screening of a large population of patients to identify those with arrhythmia and ischemia to aid the cardiologist in diagnosing these diseases is one of the biggest challenges faced by the scientific community. Chapter 5 contains a discussion of an integrated system that can identify cardiac diseases using the Gaussian mixture model (GMM) and electrocardiogram (ECG) signals. This chapter explains feature extraction, classification using GMM, and classifier performance evaluation based on error bounds on classification error. The performance of GMM is compared with *k*-means and fuzzy *c*-means clustering.

Epilepsy, one of the most common neurological disorders in the world, affects more than 60 million people (approximately 1% of the world population). Nonlinear analysis has been widely used to characterize the dynamics of transitions between states that precede the onsets of seizures. Most studies focus on finding the earliest possible time at which significant changes in system dynamics may indicate an impending seizure. Reported prediction time can range from seconds to a few minutes or even a few hours, depending on the methods used and the recording locations. Chapter 6 contains an explanation of the development of feature extraction and supervised learning related to identifying seizure-related patterns.

Chapter 7 contains a method for the automatic recognition and classification of cardiac arrhythmia. Eight types of ECG signals, including normal beat data

and seven types of arrhythmic data that were extracted from the Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) arrhythmia database, have been chosen for this analysis. Preprocessing of the signals is performed using Pan–Tompkins algorithm. The fourth-order autoregressive model coefficients and spectral entropy of the samples around the QRS complex of the signals are used as features for classification. Comparisons of the performance indices of the classifiers reveal that the probabilistic neural network in arrhythmia classification outperforms the conventional multilayered feedforward neural networks. Thus, the chapter contains an investigation of these techniques, such as signal processing, feature extraction, and pattern recognition, to develop computer programs that automatically classify ECG signals.

A migraine is a complex neurological disorder, and migraine sufferers may present alterations in some hematological variables, have an increased risk of cerebrovascular diseases, show an impaired cerebral carbon dioxide autoregulation, and have genetic mutations. Genetic mutations of the *MTHFR* gene are correlated with migraines and with the increased risk of artery pathologies. Chapter 8 combines supervised and unsupervised metabonomic techniques to the classification of 677-*MTHFR* mutations in migraineurs. In this chapter, it is shown that metabonomics can be effectively applied in clinical practice. Results indicate that the overall correlation structure of complex systems in migraine pathology reaches a classification accuracy of roughly 90%. The results of this study confirm the importance of transcranial Doppler sonography in the metabolic profiling and follow-up studies of migraine patients.

Depression grading is a serious problem due to the subjectivity of the data/information. This is performed manually using various rating tools and, hence, is always prone to risks of human errors and bias. Often, the grading is performed as ranges, for example, mild-to-moderate or moderate-to-severe. Such ranges are often confusing for the doctors who are using this information to decide on management options. These ranges can be especially difficult to analyze when a doctor is attempting to choose an option that includes drug dosage. This problem is addressed in Chapter 9 using a backpropagation neural network for making a more accurate decision on the grade of a set of real-life depression data.

Traditional clustering methods based on conventional similarity measures are not suitable for time-series clustering. Clustering, or grouping of similar entities based on the similarity of their temporal profiles, involving a large data set is a daunting task in an experiment. Various clustering methods that take the temporal dimension of the data into account have been proposed for time-series gene expression data. Chapter 10 provides a review of alignment-based clustering approaches for time-series profiles. The method also employs the temporal relationships between and within the time-series profiles. This chapter contains an explanation of the performances of these alignment methods on many data sets, a comparison of recently proposed methods, and a discussion on their strengths and weaknesses.

A unique segmentation method for mining a three-dimensional (3-D) imaging biomarker to evaluate osteoarthritis (OA) is presented in Chapter 11. The proposed method addresses knee cartilage segmentation by constructing a 3-D smoothing B-spline active surface. An adaptive combination of edge-based and balloon parameters that enforces the capture range of external forces in the case of noises and occlusions caused by tissues is also introduced. A comparison between the results of the experiments using this method and previous 3-D validated snake segmentation show the accuracy and robustness of this new method. The resulting 3-D B-spline surface can also be extended to mining other imaging biomarkers.

Most mammogram-classification techniques can be based on either density or abnormality. Density-based classification techniques categorize mammograms into tissue density classes like fatty, glandular, and dense or into the breast imaging-reporting and data system (BIRADS) I–IV categories. Abnormality-based classification techniques perform categorization based on tissue abnormality into normal and abnormal; or normal, benign cancerous, and malignant cancerous. Chapter 12 contains an explanation of the density-based classification paradigms and abnormality-based classification.

Chapter 13 is a review of various techniques for the segmentation, analysis, and quantification of biofilm images. A combination of techniques for the segmentation of biofilm images through optimal multilevel thresholding algorithms and a set of clustering validity indices, including the determination of the best number of thresholds used for the segmentation process, is presented in this chapter. Clustering validity indices are used to find the correct number of thresholds. The results are validated through the Rand index and a quantification process performed in a laboratory. The automatic segmentation and quantification results shown in this chapter are comparable to those performed by an on-site expert.

Chapter 14 contains an analysis of applications of text mining in the medical field. Applications such as text summarization, text categorization, document retrieval, and information extraction are highlighted. Knowledge extraction from medical reports of the upper gastrointestinal (GI) tract endoscopy is discussed. The upper GI tract covers three organs: (1) esophagus, (2) stomach, and (3) duodenum. Endoscopy reports highlight the observations of these organs in text form. Text mining techniques are applied to discover the relationship between the observations of the organs in the upper GI tract. Text-classification applications as applied to medical reports are discussed in this chapter.

Information communication technology (ICT) applications in global health care are poised to render effective health care once some theoretical debates are solved. Today, mental health has gained immense attention worldwide through various organizational initiatives. As a consequence, telepsychiatry, a form of ICT used in mental health, is practiced in several international environments. Chapter 15 delves beyond the boundary of traditional telemedicine use and focuses on various research methodologies and computational scopes in mental health that are used for the screening, prediction, and management of diseases. This chapter uses knowledge engineering and management techniques to focus especially on mental

health and cases of suicide. In a nutshell, the chapter swings from applications of electronic health (e-health) through telepsychiatry service to medical decision support with the help of evidence-based computational research. In the chapter, a new term, "mental health informatics," is introduced. Mental health informatics is a cutting-edge research field that is still primarily unexplored due to various technical issues and challenges, which are discussed in Chapter 15.

In recent years, there has been rapid development and widespread deployment of biomedical systems, which have progressed from single-purpose island systems such as traditional X-ray machines to massively networked health-care systems. In Chapter 16, the complexity of these systems is discussed along with the systems' engineering design methodologies. These discussions include a proposal of complex biomedical systems, which work reliably, on time, and within budget. The second half of the chapter contains a sample design for an automated mental state detection system that follows system-engineering principles.

We have made an effort in this book to provide sufficient information and methodologies for data mining that is applied to biosignals, medical images, and biosystems for the benefit of researchers, professionals, practitioners, and educators of biomedical science and engineering.

Editors

Sumeet Dua, PhD, is currently an Upchurch endowed associate professor and the coordinator of IT research at Louisiana Tech University, Ruston, Louisiana. He also serves as adjunct faculty to the School of Medicine, Louisiana State University, Health Sciences Center in New Orleans, Louisiana. He received his PhD in computer science from Louisiana State University, Baton Rouge, Louisiana.

His areas of expertise include data mining, image processing and computational decision support, pattern recognition, data warehousing, biomedical informatics, and heterogeneous distributed data integration. The National Science Foundation (NSF), the National Institutes of Health (NIH), the Air Force Research Laboratory (AFRL), the Air Force Office of Sponsored Research (AFOSR), and the Louisiana Board of Regents (LA-BoR) have funded his research with over US$3 million. He frequently serves as a study section member (expert panelist) for the NIH's Center for Scientific Review and has served as a panelist for the NSF/Computing in Science and Engineering (CISE) Directorate. Dr. Dua has chaired several conference sessions in the area of data mining and bioinformatics, and is the program chair for the 5th International Conference on Information Systems, Technology, and Management (ICISTM-2011). He has given over 26 invited talks on data mining and bioinformatics at international academic and industry arenas, advised over 25 graduate theses, and currently advises several graduate students in this field. Dr. Dua is a coinventor of two issued U.S. patents, has (co)authored over 50 publications and book chapters, and is an author/editor of 4 books in data mining and bioinformatics. Dr. Dua received the Engineering and Science Foundation Award for Faculty Excellence (2006) and the Faculty Research Recognition Award (2007); he has been recognized as a Distinguished Researcher (2004–2010) by the Louisiana Biomedical Research Network (sponsored by NIH) and has won the Outstanding Poster Award at the NIH/National Cancer Institute (NCI) caBIG–NCRI Informatics Joint Conference; Biomedical

Informatics without Borders: From Collaboration to Implementation. Dr. Dua is a senior member of the IEEE Computer Society, a senior member of the Association for Computing Machinery (ACM), and a member of SPIE and the American Association for Advancement of Science.

Rajendra Acharya U., PhD, DEng, is a visiting faculty at Ngee Ann Polytechnic, Singapore; an associate faculty member at Singapore Institute of Management (SIM) University, Singapore; and an adjunct faculty at the Manipal Institute of Technology, Manipal, India. Dr. Acharya received his PhD from the National Institute of Technology, Karnataka, Surathkal, India, and DEng from the Chiba University, Japan. He has published more than 120 papers in refereed international journals (94), international conference proceedings (30), and books (9 including those that were in press at the time of this book's publication), and has an H-index of 12. Currently, Dr. Acharya is on the editorial board of the following journals: *Journal of Medical Imaging and Health Informatics*, *Open Access Medical Informatics Journal*, *Open Access Artificial Intelligence Journal*, *Journal of Biomedical Nanotechnology*, *Journal of Mechanics in Medicine and Biology*, and the *Biomedical Imaging and Intervention Journal*. He has served as the guest editor for many journals, and is a senior IEEE member and a fellow of IETE.

Contributors

Rajendra Acharya U.
Department of Electronics and Communication Engineering
Ngee Ann Polytechnic and SIM University
Singapore

Ananthakrishna T.
Department of Electronics and Communication Engineering
Manipal Institute of Technology
Manipal, India

Najib T. Ayas
Faculty of Medicine
University of British Columbia
Vancouver, British Columbia, Canada

Venkataraya P. Bhandary
Department of Psychology
KMC
Manipal, India

Conrad Burden
Centre for Bioinformation Science
Mathematical Sciences Institute and John Curtin School of Medical Research
The Australian National University
Canberra, Australia

Gerardo Carcamo
Center for Biotechnology and Faculty of Biological Sciences
University of Concepción
Concepción, Chile

Chandan Chakraborty
School of Medical Science and Technology
Indian Institute of Technology
Kharagpur, India

Subhagata Chattopadhyay
Department of Computer Science and Engineering
National Institute of Science and Technology
Berhampur, India

Matilde Chessa
Sardinian Mediterranean Imaging Research Group
Sassari, Italy

Alan Chiu
Department of Biomedical Engineering
Louisiana Tech University
Ruston, Louisiana

Nicola Culeddu
National Council of Research
Institute for Biomolecular Chemistry
Sassari, Italy

Xian Du
Computer Science Program
Louisiana Tech University
Ruston, Louisiana

Sumeet Dua
Computer Science Program
Louisiana Tech University
Ruston, Louisiana

Oliver Faust
School of Engineering
Ngee Ann Polytechnic
Singapore

Pierangela Giustetto
Department of Neuroscience
Gradenigo Hospital
Torino, Italy

Santosh D. Hajare
Department of Gastroenterology
K.L.E. Prabhakar Kore Hospital and Research Center
Belgaum, India

Haseena H.
Department of Electrical and Electronics Engineering
MES College of Engineering
Kuttippuram, India

K. Paul Joseph
Department of Electrical Engineering
National Institute of Technology
Calicut, India

Preetisha Kaur
Department of Biomedical Engineering
Manipal Institute of Technology
Manipal, India

Krzysztof Kielan
Department of Psychiatry
Diana, Princess of Wales Hospital
Grimsby, United Kingdom

Kumar K. B.
Department of Psychology
KMC
Manipal, India

Mila Kwiatkowska
Department of Computing Science
Thompson Rivers University
Kamloops, British Columbia, Canada

William Liboni
Department of Neuroscience
Gradenigo Hospital
Torino, Italy

Teik-Cheng Lim
School of Science and Technology
SIM University
Singapore

Cesare Manetti
Department of Chemistry
"La Sapienza" University
Rome, Italy

Roshan Joy Martis
School of Medical Science and Technology
Indian Institute of Technology
Kharagpur, India

Abraham T. Mathew
Department of Electrical Engineering
National Institute of Technology
Calicut, India

Filippo Molinari
Department of Electronics
Politecnico di Torino
Torino, Italy

Alioune Ngom
School of Computer Science
University of Windsor
Windsor, Ontario, Canada

Niranjan U. C.
Department of Electronics and Communication Engineering
Manipal Institute of Technology
Manipal, India

Fethi Rabhi
School of Computer Science and Engineering
The University of New South Wales
Sydney, Australia

Ajoy Kumar Ray
Department of Electronics and Electrical Communications
Indian Institute of Technology
Kharagpur, India

Dario Rojas
Department of Computer Science
University of Atacama
Copiapo, Chile

Luis Rueda
School of Computer Science
University of Windsor
Windsor, Ontario, Canada

C. Frank Ryan
Faculty of Medicine
University of British Columbia
Vancouver, British Columbia, Canada

S. S. Saraf
Research Center, Department of Electronics and Communications Engineering
Gogte Institute of Technology
Belgaum, India

Chong Wee Seong
School of Science and Technology
SIM University
Singapore

Kumara Shama
Department of Electronics and Communication Engineering
Manipal Institute of Technology
Manipal, India

Harpreet Singh
Computer Science Program
Louisiana Tech University
Ruston, Louisiana

Numanul Subhani
School of Computer Science
University of Windsor
Windsor, Ontario, Canada

G. R. Udupi
Vishwanathrao Deshpande Rural Institute of Technology
Haliyal, India

Homero Urrutia
Center for Biotechnology and Faculty of Biological Sciences
University of Concepción
Concepción, Chile

Maria Cristina Valerio
Department of Chemistry
"La Sapienza" University
Rome, Italy

Chapter 1

Feature Extraction Methods in Biomedical Signaling and Imaging

Xian Du and Sumeet Dua

Contents

1.1 Introduction to Biomedical Signaling and Imaging

Biomedical signaling addresses certain types of signals that are obtained from in vivo organisms. Each living organism consists of a variety of systems. Each system consists of organs, tissues, and cells that fulfill certain physiological functions. For example, a cardiovascular system pumps blood to deliver nutrients to the body. Each physiological function corresponds to a biological process, which has a group of characteristic signals. According to their origin and physics, these signals can be categorized into biochemical signals (e.g., hormones and neurotransmitters), electrical signals (e.g., potentials and currents), and mechanical signals (e.g., pressure and temperature). A normal biological process typically presents a group of stable signals with tolerable variations. The abnormal deviation of these signals can indicate pathology of diseases. The retrieval of normal and abnormal features from biomedical signals can help pathological analysis and decision making. For example, a high body temperature may indicate the flu, arrhythmias in an electrocardiogram (ECG) may indicate cardiovascular disorders, and peaks or valleys in an electroencephalogram (EEG) may help to diagnose sleep disorders.

Biomedical signals can be obtained quantitatively or qualitatively from a variety of biomedical detection instruments. These signals can be analog, digital, or data stream. Among biomedical signals, imaging is employed the most because it is noninvasive and observable for operations. As shown in Figure 1.1, biomedical imaging

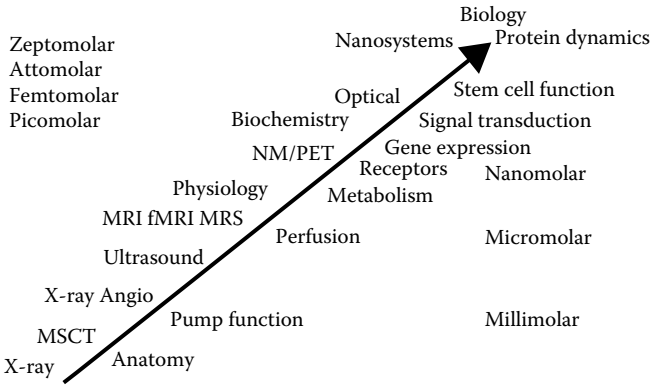

Figure 1.1 Relative sensitivity of imaging technologies. (Reprinted from *Mol Oncol* 2, Fass, L., Imaging and cancer: A review, 115–52, Copyright 2008, with permission from Elsevier.)

technologies are ubiquitous in disease measurement. The measured data are used for biomedical diagnosis and analysis after signal and image processing.

In this chapter, we first introduce biomedical signaling techniques. Second, we focus on biomedical imaging techniques that compose principal components in biomedical signaling.

1.2 Biomedical Signaling

The obtained biomedical data are of two parts: (1) signals and (2) noise. In general, a biomedical signal is defined as the useful information of the in vivo system, and noise refers to the undesired information involved in data sources. Biomedical noises originate from two sources: (1) measurement process and (2) biological system. Noises from the measurement process are caused by the limited capacity of measurement devices or software, for example, nonuniform lighting conditions cause microscopy images to have nonuniform distribution of intensity in the background. Noises from the biological system are caused by the occlusion of other uninteresting tissues in the region of interest (ROI), for example, lesions and liquids generate similar intensities with cartilage and bone in magnetic resonance imaging (MRI) images. Noises are deleted through spectral analysis or various filters.

Furthermore, we can partition the remaining signal into two parts: (1) irrelevant information and (2) information of interest. Which part constitutes the information of interest depends on specific application purposes. For example, cell images from a microscope contain cell information and information about other tissues and light. Clinical researchers are interested in cell shape for cell growth analysis, and in such studies information about light is considered as noise. In biomedical signal processing, researchers attempt to remove noises and irrelevant and redundant information in biomedical signals to facilitate the extraction of biomarkers and other characteristics of pathology for clinical or research decision making.

In accordance with different roles of signal processing, the workflow of biomedical signaling can be partitioned into five stages, as shown in Figure 1.2. In this chapter, we assume signals have been captured and preprocessed (e.g., by denoising filters) for specific feature extraction and pattern recognition. Depending on specific clinical applications and purposes, various biomedical features are selected and extracted from the preprocessed signals. In terms of applications, feature extraction in biomedical signaling is application specific, for example, minimum thickness in cartilage image segmentation. In terms of signal processing, we can understand

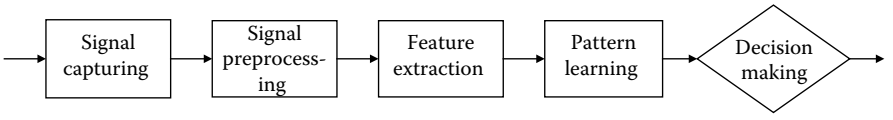

Figure 1.2 The workflow of biomedical signaling and imaging.

feature extraction topologically and statistically. For example, we can use high frequencies and low frequencies to represent noise and signals. We can use statistical information contained in signals, such as mean and standard deviation, and frequencies of certain patterns to characterize signals of interest. In this chapter, we concentrate on the feature extraction process. Pattern learning is used here to associate the extracted feature values with class labels, for example, normal and cancerous. Using the labeled sample signals, the learning model is trained until the classification result satisfies an adaptable accuracy. The obtained model can be applied to upcoming signals and can help clinicians and scientists to make diagnostic decisions.

1.3 Feature Extraction in Biomedical Signaling

In biomedical signaling, feature extraction can be conducted in two domains: frequency domain and statistical domain. In Section 1.3.1, we discuss frequency-feature extraction methods, which include Fourier transform (FT) and discrete wavelet transform (DWT).

1.3.1 Feature Extraction in the Frequency Domain

Biomedical signaling addresses two types of features, the overall shape of the spectrum and the local parameters, by using Fourier transform and discrete wavelet transform.

Let $x(t)$ denote a biomedical signal function of time t. In continuous FT, the continuous signal $x(t)$, $t \in (-\infty, +\infty)$, is decomposed into a series of sine waves with respect to the series of frequencies u, $u \in (-\infty, +\infty)$. Conversely, the original signal $x(t)$ can be obtained by using inverse FT. The pair of FT and inverse FT can be expressed as follows:

$$F(u) = \int_{-\infty}^{+\infty} x(t) e^{-j2\pi ut}\, dt \leftrightarrow x(t) = \int_{-\infty}^{+\infty} F(u) e^{j2\pi ut}\, du \tag{1.1}$$

In Equation 1.1, $j = \sqrt{-1}$ and $u \in (-\infty, +\infty)$. In the discrete form, the FT and inverse FT of the discrete variable $x(n), n \in (0, 1, 2, \ldots, N-1)$, can be expressed as follows:

$$F(u) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-j2\pi un/N} \leftrightarrow x(n) = \frac{1}{N} \sum_{n=0}^{N-1} F(u) e^{j2\pi un/N} \tag{1.2}$$

In Equation 1.2, n denotes the index of frequency members in the FT series, and N is the total number of such members. The magnitude of Fourier coefficients decreases with the increase of frequencies. The first few Fourier coefficients have more significant impacts on the signal. Hence, these coefficients are generally extracted from signals as principal features for signal analysis in the frequency domain. The FT provides no signal information in the time domain.

Wavelet transform (WT) is different from FT in that WT has a localization feature in the time domain and wavelet localization in frequency. The WT allows multiresolution analysis. For example, WT allows image description in terms of frequencies at a local position, whereas FT transforms the spatial image to a frequency domain across the image. In continuous WT, the continuous signal $x(t)$, $t \in (-\infty, +\infty)$, is discomposed into a series of wavelets in different times as follows:

$$x_w(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \varphi\left(\frac{t-b}{a}\right) \mathrm{d}t \tag{1.3}$$

In Equation 1.3, function $\varphi\left(\frac{t-b}{a}\right)$ denotes the mother wavelet, and it is continuous in both the time and frequency domains.

In the discrete form, the WT of sequence variables $x(n)$, $n \in (0,1,2,\ldots,N-1)$, can be expressed in the following orthonormal form:

$$x[n] = \sum_k \langle \varphi_k[l], x[l] \rangle \varphi_k[n] = \sum_k X[k] \varphi_k[n] \tag{1.4}$$

In Equation 1.4, $X[k]$ is the transform of $X[n]$ and $X[k] = \langle \varphi_k[l], x[l] \rangle = \sum_l \varphi_k \times [n] x[l]$. Function $\varphi_k[n]$ satisfies the following orthonormal constraint: $\langle \varphi_k[n], \varphi_l[n] \rangle = \delta[k-1]$.

Both FT and WT are broadly employed in time-series signal analysis and image processing in biomedical engineering, such as ECG and multiresolution analysis of biomedical image processing. The Gabor transform (GT) is among the popular WT approaches in image processing. One-dimensional (1-D) GT is given by

$$G_w(t) = \mathrm{e}^{-ju_0 t} \mathrm{e}^{-\left(\frac{t-t_0}{a}\right)^2} \tag{1.5}$$

Equation 1.5 can be regarded as the convolution of a Gaussian kernel function and sine waves. Thus, the most significant difference between GT and FT is that FT has no Gaussian kernel function. By setting constant a to be large or u_0 to be zero, we can obtain signal information using the frequency or spatial domain, respectively. To apply GT in a two-dimensional (2-D) image, we need to extend Equation 1.5 as follows:

$$G_w(x,y) = \frac{1}{\sigma\sqrt{\pi}} \mathrm{e}^{-j2\pi u_0((x-x_0)\cos(\theta)+(y-y_0)\sin(\theta))} \mathrm{e}^{-\left(\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}\right)} \tag{1.6}$$

In Equation 1.6, x_0 and y_0 denote the center of the Gaussian ellipse; θ denotes orientation of the 2-D Gaussian ellipse; u_0 denotes the frequency variance in radial direction, and σ denotes spatial variance of the wavelet orientation in radians.

1.3.2 Feature Extraction in the Statistical Domain

In Section 1.3.2, we discuss three statistical feature extraction methods, which include principle component analysis (PCA), independent component analysis (ICA), and information-theoretic feature extraction.

1.3.2.1 Principal Component Analysis

PCA represents raw data in a lower dimensional and combined feature space to describe the variances of the data along the extracted dimensions, expressed by eigenvectors. In PCA, principal feature components, called eigenvalues, are extracted along the new dimensions. Given a data set $\{x_1,\ldots,x_n\}$ in a d-dimensional feature space, we put these data points in a matrix X in which each row represents a data point and each column represents an attribute. Now we have the matrix $X=\left[x_1,\ldots,x_n\right]^T$, where T denotes transpose. We adjust the data points to be centered around zero by $X-\bar{X}$, where $\bar{X}$ denotes the matrix in space $\mathbb{R}^{n\times d}$, with each row representing the mean of all rows in matrix X. Such an operation ensures that the PCA result will not be skewed due to the difference between features. Then, an empirical covariance matrix of $X-\bar{X}$ can be obtained by $C=1/d\sum(X-\bar{X})(X-\bar{X})^T$. Next, we obtain a matrix V, $V=\left[v_1,\ldots,v_d\right]$, of eigenvectors in space $\mathbb{R}^d$, which consists of a set of d principal components in d dimensions. Each eigenvector v_i, $i=1,\ldots,m$, in matrix V corresponds to an eigenvalue $\lambda_i$in the diagonal matrix D, where $D=V^{-1}CV$ and $D_{ij}=\lambda_i$, $i=j$; otherwise, $D_{ij}=0$. Finally, we rank eigenvalues and reorganize the corresponding eigenvectors such that we can find the significance of variances along eigenvectors. Mathematically, we can represent the ith principal component and eigenvector v_i as follows:

$$v_i=\arg\max_{|v|=1}\left\|\left((X-\bar{X})-\sum_{k=1}^{i-1}(X-\bar{X})v_kv_k^T\right)v\right\| \tag{1.7}$$

In Equation 1.7, $(X-\bar{X})v_j$ captures λ_i, the amount of variance projected along v_j. When $i=1$, $v_1=\arg\max_{|v|=1}\left\|(X-\bar{X})v\right\|$. The PCA algorithm includes the following four steps:

Step 1: Subtract the data mean in all dimensions to produce a data set with zero mean.
Step 2: Calculate the covariance matrix.
Step 3: Calculate the eigenvectors and eigenvalues of the covariance matrix.
Step 4: List the eigenvectors according to the ranks of the eigenvalues from the highest to the lowest.

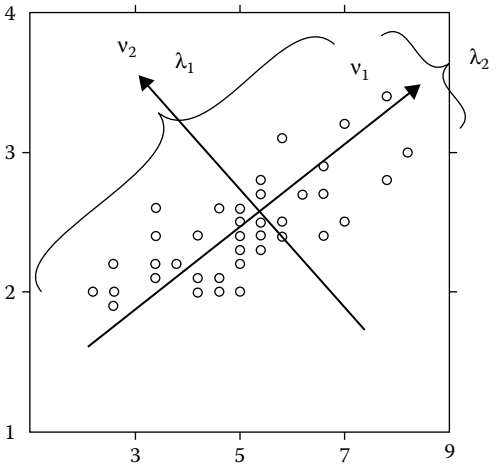

Figure 1.3 Example of principal component analysis application in a two-dimensional Gaussian mixture data set.

As shown in Figure 1.3, given a 2-D Gaussian mixture data set, v_1 and v_2 are the first and second principal components obtained by PCA. The values λ_1 and λ_2 are the corresponding first and second eigenvalues. The principal components are orthogonal in the feature space, whereas v_1 represents the most significant component of original variance in the data set, and v_2 explains the second most significant component of the remaining variance.

1.3.2.2 Independent Component Analysis

ICA attempts to discover the independent source signals from a set of observations. To simplify the expression, we reuse the data set X described in Section 1.3.2.1. Assume that k signal sources generate k signals $\{s_1,\ldots,s_k\}$ independently and that each signal has n dimensions. We obtain a $k \times n$ matrix $S = \left[s_1,\ldots,s_k\right]^T$. Given an $d \times k$ mixture matrix A, we obtain the following equation:

$$X^T = AS \tag{1.8}$$

Approximating the inverse mixture matrix A by $W = \hat{A}^{-1}$, we obtain

$$S = WX^T \tag{1.9}$$

In ICA algorithms, a function $f(s_1,\ldots,s_k)$ is proposed to measure the independence between signal sources. The maximization of this function solves the independence

analysis among signals. For example, the following joint probability function is commonly employed to measure the independence between observed signals:

$$P(S) = \Pi p(s_i) \tag{1.10}$$

In Equation 1.10, $p(s_i)$ is the probability intensity of signal source s_i, $i = 1,\ldots,k$. Using the maximum likelihood estimation (MLE), we can find the optimum statistic parameters to describe Equation 1.10.

1.3.3 Information-Theoretic Feature Extraction

Entropy is defined as follows:

$$H(x) = -\sum_{x \in \chi} p(x) \log_2 p(x) \tag{1.11}$$

In Equation 1.11, $p(x)$ is the probability of finding the feature x in the feature space χ given some end points.

Mutual information (MI) is well-known for representing the dependence between features, and it can be used to represent correlation in clinical practice. Minimum MI equals zero when no correlation exists between independent features. Higher values of MI occur when all features are strongly dependent in a group. Pairwise MI, the correlation between features X and Y, is represented as follows:

$$\mathrm{MI}(X, Y) = H(X) + H(Y) - H(X, Y) \tag{1.12}$$

1.4 Biomedical Imaging

Biomedical imaging is a type of noninvasive signal processing. It is ubiquitously employed in almost all phases of disease diagnosis, such as cancer biopsy guidance for detection, staging, prognosis, therapy planning, therapy guidance, therapy response, recurrence, and palliation (Fass 2008).

Biomedical imaging has the same workflow as biomedical signaling, as shown in Figure 1.2. Most biomedical image capturing techniques depend on interactive signals generated by electromagnetic radiation (e.g., MRI) with or without reflection (e.g., ultrasound) on in vivo organisms. As shown in Figure 1.4, such imaging technologies play an important role in generating various frequencies of electromagnetic radiation. The underlying physics and properties of imaging technologies cause various sensitivities, including temporal and spatial resolution sensitivities, for example, positron emission tomography (PET) and MRI have 1 nmol/kg and 10 umol/kg sensitivities, respectively. For different purposes of clinical diagnosis and analysis, various imaging technologies of various sensitivities are properly employed in biomedical engineering, such as MRI for knee osteoarthritis (OA),

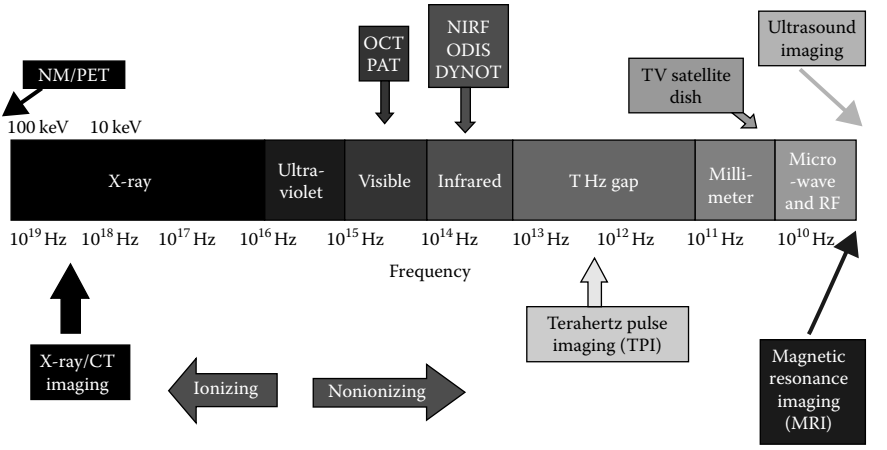

Figure 1.4 Frequency spectrum of electromagnetic radiation imaging technologies. (Reprinted from *Mol Oncol* 2, Fass, L., Imaging and cancer: A review, 115–52, Copyright 2008, with permission from Elsevier.)

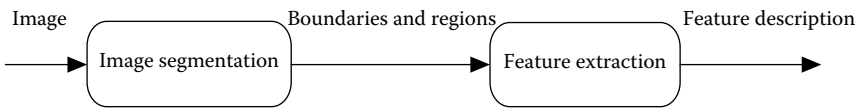

Figure 1.5 Image feature extraction.

computed tomography (CT) for mammogram, and ECG for arrhythmias. Image preprocessing consists of denoising, image enhancement, image transformation, image smoothing, and image sharpening. Image preprocessing is aimed at providing easily recognizable ROIs for image segmentation. Image segmentation refers to the process of segmenting ROIs in a preprocessed image. It constitutes the first step in feature extraction and faces the most challenges in practice.

1.4.1 Image Feature Extraction

Image feature extraction consists of two procedures: (1) image segmentation and (2) feature description. As shown in Figure 1.5, image segmentation algorithms result in boundaries and ROIs. Representative features need to be selected for machine learning. For example, boundaries generally denote the discontinuity of image intensities or abrupt changes of intensity in neighboring pixels/voxels. Each region consists of the pixels/voxels that satisfy a similarity criterion. The segmentation result of biomedical images depends on imaging modalities, intensity resolution, and anatomy quality, as these factors are the constraints of image processing techniques. For example, MRI technology can produce three-dimensional (3-D) images with around 10-umol/kg resolution. For application-specific purposes, biomedical

features can be extracted and represented from the segmented boundaries or regions. For example, geometric features of cell shape are extracted for cell growth analysis and pathological analysis of cancerous cells. Biomedical image segmentation and feature extraction varies according to medical applications and related organisms. Thus, a large number of research studies in the literature report the progress of image feature extraction in specific domains, for example, MRI image segmentation (Clarke et al. 1995; Wells et al. 1996) and brain image segmentation (Ashton et al. 1995; Zhang, Brady, and Smith 2001). A higher-level brief of biomedical imaging techniques over these application-specific topics can help readers in learning biomedical image processing horizontally in depth. Following this idea, we summarize the image segmentation methods and feature descriptions that are the most employed in biomedical engineering in Sections 1.4.2.1 through 1.4.2.5.

1.4.2 Biomedical Image Segmentation

In Section 1.4.2, we categorize biomedical image-segmentation methods into five groups, which include intensity-discontinuity-based segmentation, regional-intensity-based segmentation, hybrid algorithms combining edge-based and region-based segmentation, deformable-model-based, and pattern-classification-based segmentation.

1.4.2.1 Segmentation Based on Intensity Discontinuity

This category of segmentation has a premise that the boundary of region (structure or object) of interest shows discontinuous intensity transitions. The discontinuity from the ROI to its surroundings can be measured by the first and second derivatives of profiles across the boundaries of the ROI. The first derivatives are obtained by approximating the gradient of an image, and the second derivatives can be represented by Laplacian operators.

Let us consider a 2-D image $I(x, y)$, and (x, y) denotes the coordinates of each pixel in image $I(x, y)$. The first derivatives of the image $I(x, y)$ can be obtained by

$$\left[\frac{\partial I(x,y)}{\partial x}, \frac{\partial I(x,y)}{\partial y}\right] \tag{1.13}$$

To implement Equation 1.13 at image pixels, various masks are designed for each pixel $I(x, y)$, which is located at the center of the masks. These masks can approximate the gradient of an image by using Roberts cross-gradient, Prewitt, Sobel, or Canny operators. For example, pixel $I(x, y)$ and its neighborhood are shown in Figure 1.6, and various 3×3 masks (see Figure 1.7) can be applied on pixel $I(x, y)$. The operation of masks can be expressed as follows:

$$G(x,y) = \sum_{j=-1,0,1} \sum_{i=-1,0,1} I(x+i, y+j) w(x+i, y+j) \tag{1.14}$$

$I(x-1, y-1)$	$I(x-1, y)$	$I(x-1, y+1)$
$I(x, y-1)$	$I(x, y)$	$I(x, y+1)$
$I(x+1, y-1)$	$I(x+1, y)$	$I(x+1, y+1)$

Figure 1.6 Pixel $I(x, y)$ and its 3 × 3 neighborhood.

$w(x-1, y-1)$	$w(x-1, y)$	$w(x-1, y+1)$
$w(x, y-1)$	$w(x, y)$	$w(x, y+1)$
$w(x+1, y-1)$	$w(x+1, y)$	$w(x+1, y+1)$

Figure 1.7 The 3 × 3 mask of pixel $I(x, y)$.

−1	−1	−1
−1	8	−1
−1	−1	−1

Figure 1.8 Template of Laplacian.

The Laplacian operator of image $I(x, y)$ is defined as

$$\nabla^2 I(x, y) = \left[\frac{\partial I^2(x, y)}{\partial x^2}, \frac{\partial I^2(x, y)}{\partial y^2} \right] \tag{1.15}$$

To implement the Laplacian in Equation 1.15 at image pixels, approximation masks can be obtained, for example, the template in Figure 1.8 (Gonzalez and Woods 2002). This form of the Laplacian generates problems of unacceptable sensitivity to noise and double edges. In practice, the Laplacian is employed after convolving the following Gaussian function with an image $I(x, y)$:

$$G(x, y) = -\exp(-0.5 \times (x^2 + y^2)/\sigma^2) \tag{1.16}$$

The convolution can be further obtained by

$$\nabla^2 (G(x, y) \times I(x, y)) = \nabla^2 G(x, y) \tag{1.17}$$

In Equation 1.17, $\nabla^2 G(x, y)$ denotes the Laplacian of a Gaussian (LoG), which can be obtained by

$$\nabla^2 G(x, y) = -((x^2 + y^2)/\sigma^4 - 2/\sigma^2)\exp(-0.5 \times (x^2 + y^2)/\sigma^2 \tag{1.18}$$

The other popular segmentation techniques include edge linking by Hough transform (Kittler and Illingworth 1988), graph-based techniques (Malik and Jitendra 2000), and Canny's approach (Canny 1986).

1.4.2.2 Segmentation Based on Regional Intensity Similarity

This category of segmentation has a premise that the region (structure or object) of interest has the same intensity distribution inside, although different intensity distributions exist across regions. Thresholding has a premise that the segmented objects distribute distinctively in histograms. Threshold segmentation is a method that separates an image into a number of meaningful regions by selecting threshold values, such as intensity values. If the image is a gray image, thresholds are integers in the range $[0, L-1]$, where $L-1$ is the maximum intensity value. Normally, this method is used to segment an image into two regions, background and object, with one threshold. The following is the equation for threshold segmentation:

$$I_B(x,y) = \begin{cases} 1, & \text{if } I(x,y) > T \\ 0, & \text{if } I(x,y) \le T \end{cases} \tag{1.19}$$

In this equation, I_B is the segmentation resultant. The most famous threshold method was proposed by Otsu (1979). Otsu's method finds the optimal threshold T among all the intensity values from 0 to $L-1$ and chooses the value that produces the minimum within-class variance σ^2_{within} as the optimal threshold value. Consequently, the optimal value of T, T_{opt}, is obtained by the following optimal computation:

$$\sigma^2_{\text{within}}(T_{\text{opt}}) = \min_{0 \le T \le L-1} \left[\sigma^2_{\text{within}}(T)\right] \tag{1.20}$$

In the whole image, variances σ^2 are composed of two parts: $\sigma^2 = \sigma^2_{\text{within}}(T) + \sigma^2_{\text{between}}(T)$. Otsu shows that $\min_{0 \le T \le L-1} \left[\sigma^2_{\text{within}}(T)\right]$ is the same as $\max_{0 \le T \le L-1} \left[\sigma^2_{\text{between}}(T)\right]$. Therefore, the optimal value of T can also be obtained using the following alternative optimization process:

$$\sigma^2_{\text{between}}(T_{\text{opt}}) = \max_{0 \le T \le L-1} \left[\sigma^2_{\text{between}}(T)\right] \tag{1.21}$$

Equation 1.21 is often used to find the optimal threshold value for simple calculations. Theoretically, $\sigma^2_{\text{between}}(T)$ is expressed in the following equation:

$$\sigma^2_{\text{between}}(T_{\text{opt}}) = \omega_1(T)\omega_2(T)(\mu_1(T) - \mu_2(T))^2 \tag{1.22}$$

Using an intensity histogram, the optimal threshold T is exhaustively searched among $[0, L-1]$ to meet the objective according to Equation 1.22.

1.4.2.3 Hybrid Algorithms Combining Edge-Based and Region-Based Segmentation Methods

Both edge-based and region-based segmentation methods focus on a part of image information: the boundaries of ROIs or the interior features of ROIs. To improve

segmentation accuracy, these two parts of information can be combined to incorporate a more complete feature set for segmentation. The new segmentation methods are normally hybrid edge-based segmentation and region-based segmentation algorithms. Related methods include region growing, region splitting and merging, and graph cutting. We focus on watershed algorithms, and readers can extend their understanding to other methods easily.

In watershed algorithms, image pixels consist of three groups (as shown in Figure 1.9): The first group includes the pixels that denote a local minimum in the ROI. This group of pixels can be regarded as located at the bottom of barriers. The second group of pixels is located at the catchment basins of those minimums. These pixels tend to fall to a single local minimum. The third group of pixels is located at the watershed lines and has equal likelihood of falling to two or more minimums. In watershed algorithms, the objective is to detect the location of watershed lines and use those lines as boundaries to segment the ROIs.

Watershed algorithms normally start with all the pixels that have local minimum values. The algorithms use these pixels as the basis for initial watersheds. Assuming an image has an intensity level set $1 \le k \le N$, for example, $N = 255$, these algorithms iterate in the following steps to converge: For the pixels in intensity level k,

If the pixels are nearest to only one minimum, assign the pixels to the region of that minimum.
Else if the pixels are nearest to more than one minimum, label the pixels as boundary.
Else, assign the pixels to a new region.
Iterate until pixels in all levels are labeled.

Normaly, watershed algorithms use the gradient image as input.

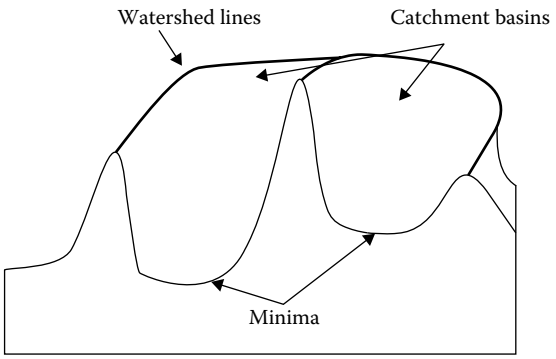

Figure 1.9 Groups of pixels/voxels in watershed algorithms.

1.4.2.4 Segmentation Based on Deformable Models

Deformable models attempt to define an evolving contour in equations such that the contour is activated toward the boundary of the ROI. According to the different representations of the contour, deformable models are categorized into two groups: (1) parametric models and (2) level-set methods. Parametric models include the active contour model (ACM; Kass, Witkin, and Terzoploulos 1988), gradient vector flow (GVF; Xu and Prince 1998b), balloon model (Cohen 1991), and active shape model (Cootes et al. 1995). Level-set methods include the level-set model by Osher and Sethian (1988), model of active contours without edges (ACWE) by Chan and Vese (2001), and Mumford and Shah's (MS) function (Jayant and Shah 1989).

1.4.2.4.1 Parametric Models

In this section, we present three popular parametric models: (1) ACM, (2) active shape model (ASM), and (3) GVF. The key difference between ACM and ASM is that "ASM can only deform to fit the data in ways consistent with the training set" (Coots et al. 1995, p. 38).

In ACM, contours can be represented as explicit or implicit parametric mathematical formulas. The contours can be approximated using polynomial expressions such as spline equations. The objective of ACM is to find the best contour, or the contour having a minimal energy. The energy of a contour is composed of two parts: (1) internal energy and (2) external energy. Kass et al. described internal energy with respect to tension (first derivative) and bending (second derivative) in the contour. The external energy is represented by the high intensity gradient. The formulation is replaced with the corresponding force balance, a vector-valued partial differential Euler–Lagrange equation. Using the gradient descent method, the desired contour can be obtained.

Traditional ACMs suffer from a number of problems: First, the initialization should be close enough to the real contour or it may lead to an unexpected result (snake has limited searching ability). The proposed solutions include multiresolution and pressure forces. Second, the internal energy term of ACMs generates a shrinking force, which implements regularization (internal continuity and smoothing) on the contour. Such a force may cause contraction of the contour, and the active contour faces difficulties in converging to a concave boundary. It is also difficult to choose a pair of proper internal parameters. The proposed solutions include balloon force (Cohen 1991) and robust active contours (Xu, Segawa, and Tsuji 1994). Third, the external forces have limited capture range. The magnitude of an external force disappears from the boundary quickly because this force is generated from the boundary. In a study by Xu and Prince (1998b), GVF has been introduced as an external force using the internal information.

Cohen (1991) introduced the balloon model to deal with the constant problem implicated in the traditional "snake" evolution: If the initialization is not close

enough to the real contour, it cannot be attracted. The balloon force is introduced into the external force as follows:

$$F = k_1 \vec{n}(s) - k \frac{\nabla I}{\|\nabla I\|} \tag{1.23}$$

where $\vec{n}(s)$ is the normal unitary vector to the curve at the point $v(s)$ and k_1 is the amplitude of the balloon force. Both k and k_1 are smaller than a pixel size. Further, k is slightly bigger than k_1 so that the edge force can stop inflation of the balloon. The numeric parameters are chosen in the same order of magnitude for elasticity and rigidity. The balloon model poses another problem in implementation. Its use raises the question of how to choose the numeric parameters. Most of the parameters have to be adjusted based on experience.

Xu et al. (1994) introduced GVF to address the problems of initialization and poor convergence to concave boundaries. The GVF is defined as a vector field $\vec{v}(x, y) = (u(x, y), v(x, y))$ that minimizes the following function:

$$\varepsilon = \iint \mu(\nabla I)\left(u_x^2 + u_y^2 + v_x^2 + v_y^2\right) + |\nabla I|^2 \, |\vec{v} - \nabla I|^2 \, \mathrm{d}x \, \mathrm{d}y \tag{1.24}$$

In Equation 1.24, $\mu(\nabla I)$ is a weighting function that is implemented to adjust the smoothing effect according to the distance between vector field $\vec{v}(x, y)$ and image gradient $\nabla I(x, y)$ (Xu and Prince 1998b).

The GVF is found by using the calculus of variations by solving the Euler equations. The external force is replaced with this GVF in the traditional snake force equation. Xu et al. (1994) applied the following weighting function in Equation 1.24:

$$u(|\nabla I|) = \mathrm{e}^{-(|\nabla I|/K)} \tag{1.25}$$

In this equation, K is a nonnegative smoothing parameter for the force field (u, v). This weighting function is a decreasing function of edge-force magnitude, and smoothes the force field only when the edge strength is low.

In ASM, contours should incorporate knowledge about the desired object in a general model. Shapes are represented by groups of landmarks. Shape models involve information of the types of objects to model, locations of the landmarks, and collection of the object images. This approach can be summarized in the following two steps: (1) training and (2) application (Cootes et al. 1995).

In the training step, we attempt to describe the available training data (clouds) using PCA and the point distribution model (PDM). Training data are chosen (dimension of the cloud or hyperellipsoid) manually with respect to the global and local characteristics in the images. The allowable shape is deduced by the mean shape and linear combination of eigenvectors.

The moving step size of each landmark is decided by the maximum edge strength (suggested or gradient) along the normal to boundary. The shape parameters

(including translation, rotation, scaling, and residual adjustments, which are adjustments of shape parameters) are adjusted according to the movements.

The training result for ASM is expressed in the form of PDM as follows: $X = \bar{X} + Pb$, where X is a shape model that consists of a number of landmarks for a 2-D image. In this equation, $\bar{X}$ is the mean shape, in the dimension of $2 \times$ the number of landmarks for the 2-D image; P is the matrix of the first several principal eigenvectors, in the dimension of $2 \times t$ for the 2-D image; b is a vector of weights, in the dimension of t; and Pb is the weighted sum of different shape-variation modes.

Here, we describe how the search method works (see Figure 1.10). It includes three steps:

Step 1: For a given point, consider a sample of k pixels along a profile on either sides of the model point (see Figure 1.10a). Put the derivative of those $2k + 1$ samples in a vector g_i. Normalize the sample by dividing the sum of absolute element values.

Step 2: Assume the sample points are distributed as a multivariate Gaussian. A criterion for evaluating the fitting of a new sample to the model is represented by Mahalanobis distance (see Figure 1.10b) as follows:

$$f(g_i) = (g_s - \hat{g})^T S_g (g_s - \hat{g}) \tag{1.26}$$

where g_s is the set of normalized samples for a given model point, $\hat{g}$ is the mean of the set of sample points, and S_g is the covariance of the set of sample points.

Step 3: Sample more pixels along the profiles ($m > k$), and choose the point with the lowest value of $f(g_s)$ as the moving destination for the model point (see Figure 1.10c).

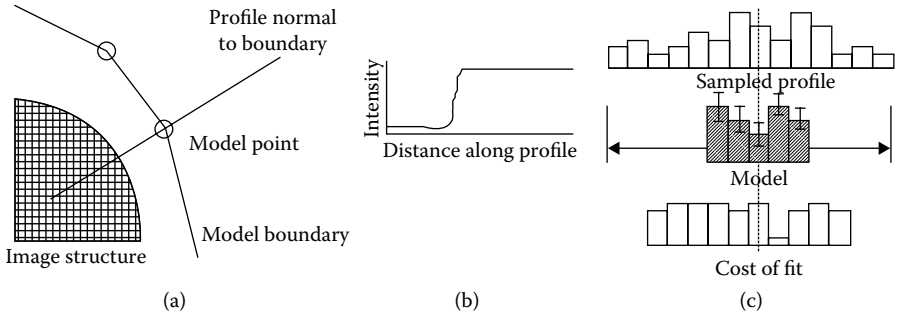

Figure 1.10 Intensity distribution: Parts (a) and (b) show the intensity distribution along a profile normal to the image boundary through a model point. Part (c) shows a statistical model for searching the fitful image boundary. (From Cootes, T. F. and C. J. Taylor, *Proc. SPIE Medical Imaging*, 2001. With permission.)

The difficulty of ASM lies in how to choose the location of training data, which determines the dimension of the ellipsoid. If any important point is missed, the final shape may definitely not be robust.

1.4.2.4.2 Level-Set Methods

Osher and Sethian (1988) introduced the level-set function to active contours in the 1980s. Zero level set represents the active contour by a level-set function $\phi(x, y)$ as follows:

$$C(t) = \{(x, y) \in I \,|\, \phi(x, y, t) = 0\} \tag{1.27}$$

Here, $\phi(x, y, t)$ denotes a Lipschitz continuous function:

$$\begin{cases} \phi(x, y, t) > 0, & \text{if } (x, y) \text{ locates inside contour } C(t) \\ \phi(x, y, t) = 0, & \text{if } (x, y) \text{ locates on contour } C(t) \\ \phi(x, y, t) < 0, & \text{if } (x, y) \text{ locates outside of contour } C(t) \end{cases} \tag{1.28}$$

According to the nonlinear partial directive equation, the evolution of an active contour can be represented as follows:

$$\frac{\partial \phi}{\partial t} = v_N(\phi) \cdot |\nabla \phi| \tag{1.29}$$

Equation 1.29 indicates that contour ϕ evolves with the velocity $v_N(\phi)$ normal to the contour.

Chan and Vese (2001) introduced the ACWE model. The ACWE model addresses the difficulty of applying boundary-based segmentation techniques to objects whose boundary is either smooth or cannot be defined by gradient. Mathematically, image segmentation is defined as a minimization problem in the ACWE model by

$$\min_{\phi, c_1, c_2} E(\phi, c_1, c_2, \lambda) \tag{1.30}$$

Here, $E(\phi, c_1, c_2, \lambda)$ denotes the energy defined by

$$\begin{aligned} E(\phi, c_1, c_2, \lambda) = {} & \int_\Omega |\nabla H(\phi)| \mathrm{d}x + \lambda \int_\Omega \big(H(\phi)(c_1 - I(x))^2 \\ & + H(-\phi)(c_2 - I(x))^2 \big) \mathrm{d}x \end{aligned} \tag{1.31}$$

where Ω denotes the image domain and H denotes the Heaviside function.

1.4.2.5 Segmentation Based on Pattern Classification Methods

We include the application of pattern classification methods in image segmentation in Chapter 2.

1.4.3 Imaging Feature Description

We categorize imaging features into low-level and high-level features. Low-level features can be extracted by processing images locally without involving any spatial integration. Low-level imaging features include pixel values, gradients, edges, corners, curvatures, histograms, colors, and statistical information such as means and variance. We can easily segment these features in biomedical images by using the intensity distribution in the images. For example, an edge-based segmentation algorithm obtains both local and global edge and gradient features in images. High-level features can be extracted by globally integrating pixel/voxel information in an image and describing regional signatures, such as shape, features, and texture. We illustrate some of the most commonly employed feature descriptors in biomedical imaging in Sections 1.4.3.1 and 1.4.3.2.

1.4.3.1 Shape Features

We categorize shape features into two groups: (1) geometric shape features and (2) topological shape features. Given a closed contour $c(\varphi) = [x(\varphi), y(\varphi)]$, we can describe the shape features of the contour, also called "shape descriptors," using $d(\varphi)$, which is the radial distance from the points on the object boundary to the center of mass of the object (reference point). A shape descriptor is an operator applied to the binary image of a shape, resulting in a numerical quantity. Shape descriptors can be region or boundary based. For example, we can obtain the object "area" A and "perimeter" P as follows:

$$P = \int_0^{2\pi} \sqrt{c(\varphi)^2 + \left(\frac{\mathrm{d}c(\varphi)}{\mathrm{d}\varphi}\right)^2} \tag{1.32}$$

$$A = \frac{1}{2}\int_0^{2\pi} c^2(\varphi)\,\mathrm{d}\varphi \tag{1.33}$$

In this section, we introduce four popular shape features in biomedical imaging. These features have common advantages in measuring shape: the features are dimensionless and invariant to rotation, translation, and reflection. First, we define "circularity" C_r as a number that measures deviations of radial distance in a shape:

$$C_r = \frac{u_R}{\sigma_R}, \quad 0 \le C_r \tag{1.34}$$

Here, u_R is the mean value of shape radial distance, and σ_R is the standard deviation of those distances. Shape "roughness" R calculates the number of angles in which more than one boundary point is observed, and divides this number by the total number of angles. Low-order moments can also calculate roughness. However, roughness is dependent on a digitalization step for computation. Shape "compactness" C_p describes the diffused nature of a brain tumor compared to a circle, which is defined as follows:

$$C_P = 4\pi \frac{A}{P^2}, \quad 0 \le C \le 1 \tag{1.35}$$

Topological shape features include connectivity, number of holes, Euler number, convex hull, skeleton, and object counting. For further information, we suggest the work by Gonzalez and Woods (2002).

1.4.3.2 Texture Feature Description

Texture features are extracted on a gray-tone spatial-dependence matrix. In this matrix, each component $M(i, j)$ represents the occurrence of paired pixels that have a specified distance d along a specified direction and combined gray levels of i and j. As shown in Figure 1.11, four directions are used in Harrilick's (Harralick, Shanmugam, and Dinstein 1973) texture feature calculation: 0°, 45°, 90°, and 135°.

We illustrate three representative Harrilick's features in Equations 1.36 through 1.38, to measure homogeneity, contrast, and gray-tone linear dependencies in images. In these equations, c_{nor} denotes the normalization constant, L denotes the gray level in image I, and μ_x, μ_y, σ_x, and σ_y denote the means and standard deviations of $M(i, j)$ along the coordinates x and y:

$$f_{\text{homogeneity}} = \sum_{i=1}^{L}\sum_{j=1}^{L}\left(\frac{M(i,j)}{C_{\text{nor}}}\right) \tag{1.36}$$

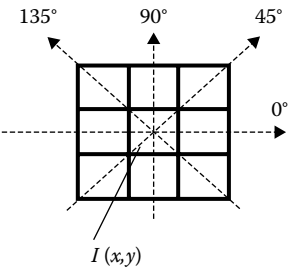

Figure 1.11 Specified direction and neighborhood of an image pixel.

$$f_{\text{contrast}} = \sum_{n=0}^{L-1} n^2 \left(\sum_{|i-j|=n}^{L} \left(\frac{M(i,j)}{C_{\text{nor}}} \right)^2 \right) \tag{1.37}$$

$$f_{\text{linearity}} = \frac{1}{\sigma_x \sigma_y} \left(\sum_{i=0}^{L} \sum_{j=0}^{L} \left(\frac{ijM(i,j)}{C_{\text{nor}}} \right) - \mu_x \mu_y \right) \tag{1.38}$$

The feature $f_{\text{homogeneity}}$ can capture significant dominant gray-tone transitions in an image. A homogeneous image results in few significant values of $f_{\text{homogeneity}}$. The feature f_{contrast} can capture local gray-tone variations in an image. The feature $f_{\text{linearity}}$ can capture linear dependent and correlative structures of gray tone along different directions in an image. Generally, additive noises result in the reduction of linear correlation between structures in an image. Harralick, Shanmugam, and Dinstein (1973) also suggests other texture features that can be extracted from a gray-tone spatial-dependencies matrix, such as variance, entropy, and information measures of correlation. Correlations or dependencies exist among the texture features when applied in image analysis. Thus, the feature selection procedure is generally employed in feature extraction. Readers are referred to the work by Huang and Thomas (1999) for details regarding the application of feature selection techniques in image retrieval.

1.5 Summary

In this chapter, we introduce the fundamental feature extraction techniques in biomedical signaling and imaging. This chapter is not a survey, and we do not review all the literature in this domain.

We categorize feature extraction techniques in biomedical signaling into frequency-based, statistics-based, and informatics-based techniques. These techniques can also be applied to biomedical imaging. Frequency-based feature extraction is used to decompose time-series signals or images into frequency components for further processing and analysis. Statistics-based feature extraction can extract principal or independent components according to a variance distribution or a dependency estimate between the source and the signals. Informatics-based feature extraction uses entropy, MI, and other related techniques to describe the significance of a feature or a subset of features.

We categorized feature extraction techniques for biomedical imaging into image segmentation and image feature descriptors. We classified biomedical image segmentation techniques into five groups: (1) boundary-based, (2) region-based, (3) integration of boundary-based and region-based, (4) deformable model–based, and (5) pattern recognition–based groups. Boundary-based segmentation attempts to find discontinuous intensity pixels for the ROI, and the related techniques most commonly use gradient to find edges and boundaries. Region-based

segmentation methods search for regions with more similar pixel intensities inside the ROI than pixel intensities across the ROI. Integration of the aforementioned two technologies can regularize the weights of both boundary-based and region-based techniques to combat difficult segmentation when the ROI is contaminated highly by noises, artifacts, and other tissues. Deformable model–based techniques are popular in biomedical imaging because these techniques are adaptive to biomedical images, especially for soft organisms. The "active" nature of these models highlights the extraction of complex and evolving features in biomedical images. For a discussion of pattern recognition–based image segmentation, readers are referred to Chapter 3. We briefly introduce a number of geometric shape feature descriptors and texture feature descriptors. Readers can formulate specific feature descriptors for various application purposes after extracting the ROIs or boundaries.

In biomedical signaling and imaging, features are sensitive and application specific. We do not include all biomedical features in this chapter. However, all the features and feature extraction techniques discussed in this chapter are applicable to any types of biomedical signals and images. The features extracted are applied in machine learning methods for clinical pattern recognition and decision making.

References

Ashton, E. A., M. J. Berg, K. J. Parker, J. Weisberg, C. W. Chen, and L. Ketonen. 1995. Segmentation and feature extraction techniques, with applications to MRI head studies. *Magn Reson Med* 33:670–7.

Canny, J. 1986. A computational approach to edge detection. *IEEE Trans Pattern Anal Mach Intell PAMI* 8:679–98.

Chan, T., and L. Vese. 2001. Active contours without edges. *IEEE Trans Image Process* 10:266–77.

Clarke, L. P., R. P. Velthuizen, M. A. Camacho, J. J. Heine, M. Vaidyanathan, L. O. Hall, R. W. Thatcher, and M. L. Silbiger. 1995. MRI segmentation: Methods and applications. *Magn Reson Imaging* 13:343–68.

Cootes, T. F., C. J. Taylor, D. H. Cooper, and J. Graham. 1995. Active shape models: Their training and application. *Comput Vis Image Underst* 61:38–59.

Cohen, L. 1991. On active contour models and balloons. *CVGIP Image Underst* 53:211–8.

Fass, L. 2008. Imaging and cancer: A review. *Mol Oncol* 2:115–52.

Gonzalez, R. C., and R. E. Woods. 2002. *Digital Image Processing.* 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Harralick, R.-M., K. Shanmugam, and I. Dinstein. 1973. Texture features for image classification. *IEEE Trans Syst Man Cybern* SMC-3:610–21.

Huang, Y. R., and S. Thomas. 1999. Image retrieval: Current techniques, promising directions, and open issues. *J Vis Commun Image Represent* 10:39–62.

Jayant, M., and D. Shah. 1989. Optimal approximations by piecewise smooth functions and associated variational problems. *Commun Pure Appl Math* 42:577–685.

Kass, M., A. Witkin, and D. Terzoploulos. 1988. Snakes: Active contour models. *Int J Comput Vis* 1:321–31.

Kittler, J., and J. Illingworth. 1988. A survey of the hough transform. *Comput Vis Graph Image Process* 44:87–116.

Malik, J., and S. Jitendra. 2000. Normalized cuts and image segmentation. *IEEE Trans Pattern Anal Mach Intell* 22:888–905.

Otsu, N. 1979. A threshold selection method from gray-level histogram. *IEEE Trans Syst Man Cybern* SMC-9 1:62–6.

Osher, S., and J. Sethian. 1988. Fronts propagating with curvature dependent speed: Algorithms based on Hamilton–Jacobi formulations. *J Comput Phys* 79:12–49.

Wells, W. M., W. L. Grimson, R. Kininis, and F. A. Jolesz. 1996. Adaptive segmentation of MRI data. *IEEE Trans Med Imaging* 15:429–42.

Xu, C., and J. L. Prince. 1998a. Generalized gradient vector flow external forces for active contours. *Signal Process* 71:131–9.

Xu, C., and J. L. Prince. 1998b. Snakes, shapes, and gradient vector flow. *IEEE Trans Image Process* 7:359–69.

Xu, G., E. Segawa, and S. Tsuji. 1994. Robust active contours with insensitive parameters. *Pattern Recognit* 27:879–84.

Zhang, Y., M. Brady, and S. Smith. 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* 20:45–57.

Chapter 2

Supervised and Unsupervised Learning Methods in Biomedical Signaling and Imaging

Xian Du and Sumeet Dua

Contents

2.1 Introduction

As we describe in Chapter 1, the objective of biomedical signaling and imaging is to detect the significant features in data sets and help clinicians or scientists recognize pathological evolution in order to better track the development of diseases and further make decisions about how to treat those diseases. As shown in Chapter 1, Figure 1.2, pattern learning methods are designed to partition the given signals and images into classes, such as those indicating healthy patients and those indicating cancerous patients. The feature extraction methods presented in Chapter 1 can provide features to compose input values of pattern learning models. Using pattern learning procedures, we attempt to discover associative features and to associate patterns of feature subsets with class labels. This learning procedure consists of training, testing, and performance evaluation (as shown in Figure 2.1). Given a pattern learning model that satisfies some performance criteria, we can apply the trained model to upcoming data from the same space and generate labels for decision making. In other words, we employ machine learning methods in biomedical signaling and imaging to facilitate the analysis, diagnosis, prognosis, and treatment of a disease or pathology, such as biomedical image segmentation and content-based retrieval of information from biomedical imaging. In biomedical signaling and imaging, machine learning performs signal and image analysis, interpretation of extracted features using semantic concepts in specific applications, and decision making on the disease or pathology. For example, given microscopy cell imaging

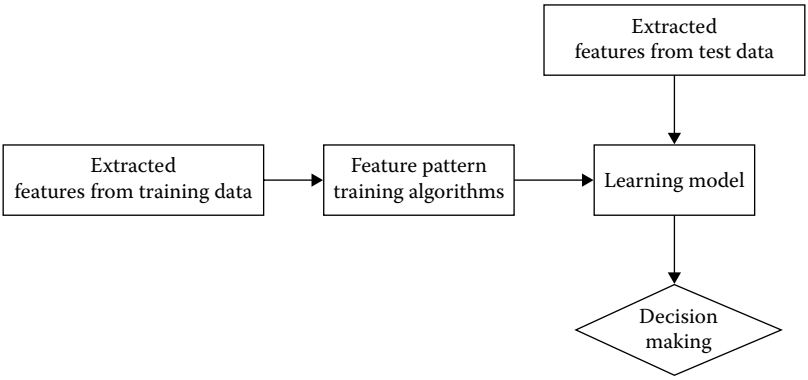

Figure 2.1 The workflow of pattern learning procedure.

and its segmentation results (as shown in Figure 2.2), we can extract cell shape features from normal cells and cancerous cells, and use these features as input to train a machine learning model. The trained model can be employed for the classification and prediction of healthy persons and cancerous patients.

As machine learning has been evaluated and employed in a large number of specific biomedical applications, readers can find related literature reviews on various topics, for example, machine learning for biomedical image segmentation (Bezdek, Hall, and Clarke 1993; Kapetanovic, Rosenfeld, and Izmirlian 2004) and for detection and diagnosis of diseases (Sajda 2006). We focus on the introduction and explanation of fundamental machine learning methods so that readers understand why and how these techniques can be employed in specific biomedical signaling and imaging domains.

In this chapter, we first introduce the fundamentals of machine learning in Section 2.2, as well as supervised learning and unsupervised learning. Then, we briefly elucidate a variety of classic supervised learning methods and unsupervised learning methods, respectively, in Sections 2.3 and 2.4. Next, we discuss a number of performance evaluation measures for machine learning methods. Finally, we describe the challenges that must be faced in biomedical signaling and imaging when we apply supervised and unsupervised learning methods. Furthermore, we explore potential research directions in the application of machine learning methods to biomedical signaling and imaging.

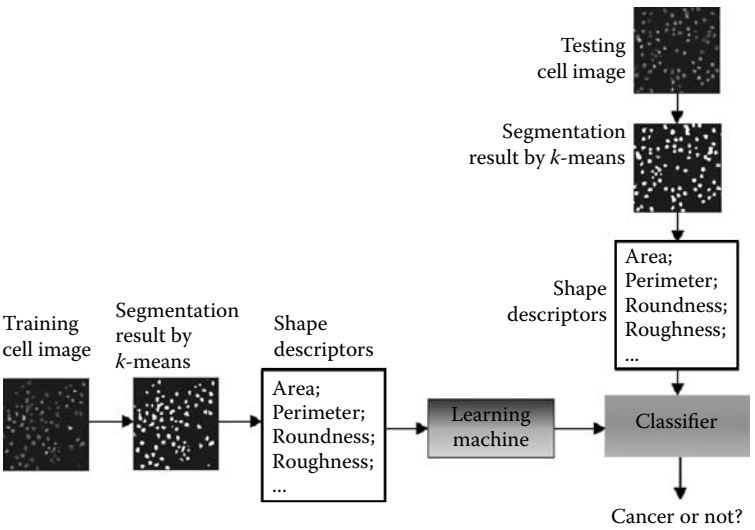

Figure 2.2 Example of machine learning for biomedical signaling and imaging: The *k*-means segmentation result of nucleus images and shape features for classification or prediction are shown. (Adapted from http://www.cellprofiler.org/)

2.2 Machine Learning

Machine learning is a computational process of automatically inferring and generalizing a learning model from sample data. Learning models most often employ statistical functions or rules to describe the dependencies among data and causalities and correlations between input and output. Theoretically, given an observed data set S, a set of parameters θ and variable x, and a learning model $f(x, \theta)$, a machine learning method aims to minimize the learning errors $E(f(x, \theta), X)$ between the learning model $f(x, \theta)$ and the ground truth. Without loss of generalization, we obtain the learning errors using the difference between the predicted output $f(x, \hat{\theta})$ and the observed sample data, where $\hat{\theta}$ is the set of approximated parameters derived from optimization procedures for the minimization of the objective function of learning errors. Machine learning methods differ from each other mainly in the selection of learning model $f(x, \theta)$, parameters θ, and the expression of learning error $E(f(x, \theta), X)$.

There are two types of machine learning methods: (1) supervised learning and (2) unsupervised learning. In supervised learning, data labels are given to training data, whereas in unsupervised learning data labels are not given to training data. In supervised learning, training data are labeled by experts. A supervised learning algorithm uses the labeled samples for training and generalizes a model structure for upcoming data points. The objective of using a supervised learning algorithm is to obtain the highest accuracy of labeling. Typical supervised learning methods include decision trees, Bayesian networks (BNs), artificial neural networks (ANNs), support vector machines (SVMs), and so on. Supervised learning takes advantage of prior knowledge and experience gathered by experts. The appropriate selection of learning structures and accurate labels can always lead to confident labels of upcoming data points that lie in the same data space as the training data. The accuracy of supervised learning can be quantitatively evaluated by various metrics, such as classification accuracy, sensitivity, and specificity. However, supervised learning relies on an expert's accuracy in labeling. No expert can guarantee that he or she is always able to label biomedical signals or images with 100% accuracy. We can employ voting among experts to solve the discrepancy problem; but voting also cannot guarantee 100% accuracy.

In order to reduce sensitivity to prior labels, unsupervised learning algorithms generalize the model structure from unlabeled data, for partitioning new data without human interference. Data are clustered based on similarity measure, density distribution, association or correlation, and other metrics. Typical unsupervised learning methods include k-means clustering, hierarchical clustering, density-based clustering, grid-based clustering, and self-organizing map (SOM) ANN (Han and Kamber 2006).

2.3 Supervised Learning Methods in Medical Signaling and Imaging

In supervised learning, pairs of input and target output are given to train a function, and a learning model is trained such that the value of the function can be predicted at a minimum cost. Based on the structure of learning algorithms according to the different objective functions, we introduce several supervised learning methods including SVM, ANN, decisions trees, Bayesian Network (BN), and Hidden Markov Model (HMM).

2.3.1 Support Vector Machines

Given a data set X in an n-dimensional feature space, SVM separates the data points in X with an $n-1$-dimensional hyperplane. In SVM, the objective is to classify the data points with the hyperplane that has the maximum distance to the nearest data point on each side. Subsequently, such a linear classifier is also called the "maximum margin classifier." As shown in Figure 2.3a, any hyperplane can be written as the set of points X satisfying $w^T x + b = 0$, where the vector w is a normal vector perpendicular to the hyperplane and b is the offset of the hyperplane $w^T x + b = 0$ from the original point along the direction of w. Given the labels of data points X for two classes, class 1 and class 2, we present the labels as $Y = +1$ and $Y = -1$. Meanwhile, for a pair of (w^T, b), we classify data X into class 1 or class 2 according to the sign of the function $f(x) = \text{sign}\,(w^T x + b)$ as shown in Figure 2.3. Thus, the linear separability of the data X in these two classes can be expressed in the combinational equation form as $y \cdot (w^T x + b) \geq 1$. In addition, the distance from a data point to the separator hyperplane $w^T x + b = 0$ can be computed as $r = \dfrac{w^T x + b}{\| w \|}$, and the data

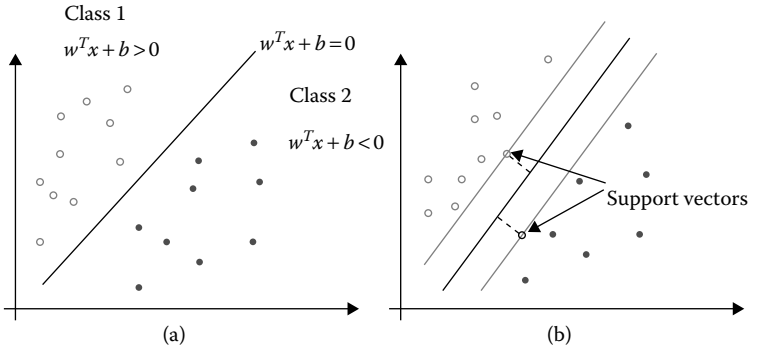

Figure 2.3 Support vector machine classification.