1 INTRODUCTION

1.1 BASIC CONCEPTS IN PATTERN RECOGNITION

Pattern recognition is characteristic to all living organisms. However, different creatures recognize differently. If a human would recognize another human by sight, by voice or by handwriting, a dog may recognize a human or other animal by smell thirty yards away which most humans are incapable of doing. Yet most dogs are unimpressed by looking at the mirror since they do not actually *recognize* another dog over there. A blind person would recognize various items just by touching them. But recognition is not restricted to objects that can be identified using biological senses. In a conversation we can suddenly identify an old argument that we heard years ago. All of these examples are classified as *recognition*.

The object which is inspected for the "recognition" process is called a *pattern*. Usually we refer to a pattern as a *description* of an object which we want to recognize. In this text we are interested in spatial patterns like humans, apples, fingerprints, electrocardiograms or chromosomes. In most cases a pattern recognition problem is a problem of discriminating between different populations. For example we may be interested among a thousand humans to discriminate between four different types: (a) tall and thin (b) tall and fat (c) short and thin (d) short and fat. We thus want to *classify* each person in one of four populations. The recognition process thus turns into *classification*. To determine which *class* the person belongs to, we must first find which *features* are going to determine this classification. The age of the person is clearly not a feature in this case. A reasonable choice of course is the pair of numbers (height, weight) and we thus perform a *feature selection* for this particular problem. Getting these measurements is called *feature extraction.*

At times feature selection may be an easy task while feature extraction is too costly. In this case we may look for an alternative way of selecting features or go ahead and extract the features of the original selection. It is not recommended to start compromising and choose less adequate features which are easier to extract. Suppose for example that a certain medical test, very expensive, is necessary to determine (together with other tests) whether a patient has some severe disease. No competent doctor would even consider dropping that test in order to 'simplify' the feature extraction.

Forecasting the weather is based upon inspecting a weather map. The map itself is raw input data on which we perform *preprocessing*. The features to look for are usually known to the professional due to vast experience. The preprocessing here includes extracting these features and identify *noise*. For an expert, one glance at a weather map is enough to produce a reasonable weather forecast. The expert knows what features to look for and if extracting them is not complicated, forecasting is straightforward. In general we may insert the knowledge acquired by experts in this field into an expert system that will replace the expert and will (almost) always provide a good weather forecast.

Medical diagnosis is another example of a pattern recognition problem where feature selection is a very delicate process since quite often human life is in stake. The features are usually some test results like blood pressure or blood sugar rate, or symptoms like 'coughing at night' or not having feeling in the forefinger'. Features of a completely different nature are 'no heart problem in the family' or 'the patient had already this disease'. An appropriate feature extraction in medical diagnosis rely first on the objective test results and then on the patient's ability to provide an accurate description of the symptoms and 'related facts' in his family history. It is more than relevant for example for a person who is treated for hearing loss, to mention whether there are deaf people in his family.

In designing a *pattern recognition system*, i.e. a system that will be able to obtain an unknown incoming pattern and classify it in one (or more) of several given classes, we clearly want to employ all the available related information that was previously accumulated. We assume that some *sample patterns* with known classification are available. These patterns with their typical attributes form a *training set* which provides relevant information how to associate between input data and *decision making*. By using the training set the pattern recognition system may learn various types of information like statistical parameters, relevant features, etc.

The dominant concept in pattern recognition is that of *clustering*. A *cluster* consists of a number of *similar* objects (patterns) which are grouped together. We may consider a cluster of points in the n-dimensional space, a cluster of stars which seem to be grouped together or a cluster of people in the community whose annual income is under \$20,000 per year. If we consider a cluster of people with 'low' income we take a further step and define a *fuzzy cluster*. Clustering given input data is a major subject in pattern recognition. It consists of dividing the data into clusters and establishing the *cluster centers* and *cluster boundaries*. An *a priori* knowledge of the number of clusters and their approximate locations definitely simplifies our task. We then carry a *supervised learning* process. If the data is of no known characteristic we obtain an *unsupervised learning* process.

Given input data it can be clustered in several ways. For example let the input consist of all the schools in town. If we cluster them geographically we get one set of clusters. If on the other hand we find *similarity* between schools only if the number of their students is similar, we obtain a different set of clusters. If we consider the attribute 'quality' we obtain a third set of clusters and this last partition is even ambiguous since people measure 'quality' differently.

1.2 CLASSIFIERS

The final goal in pattern recognition is classification of a pattern. From the original information that we obtain about the pattern we first identify the relevant features and then use a *feature extractor* to measure them. These measurements are then passed to a *classifier* which performs the actual classification, i.e., determines at which of the existing classes to classify the pattern. If the pattern is for example 'noise' it is rejected by the classifier.

In this section we assume the existence of *natural grouping*, i.e. we have some *a priori* knowledge about the classes and the data. For example we may know the exact or approximate number of the classes and the correct classification of some given patterns which are called the *training patterns*. Usually, it is this type of information and the type of the features that may suggest which classifier to apply for a given application.

Decision Functions

When the number of classes is known and when the training patterns are such that there is geometrical separation between the classes we can often use a set of decision functions to classify an unknown pattern. Consider for example a case where two classes C_1 and C_2 exist in R^n and a hyperplane $d(\boldsymbol{x})=0$ which separates between their patterns can be found. Then we can use the *decision function* $d(\boldsymbol{x})$ as a *linear classifier* and classify each new pattern by

$$\begin{aligned} d(\boldsymbol{x}) > 0 &\Rightarrow \boldsymbol{x} \in C_1 \\ d(\boldsymbol{x}) < 0 &\Rightarrow \boldsymbol{x} \in C_2 \end{aligned} \tag{1.2.1}$$

The hyperplane $d(\boldsymbol{x})=0$ is called a *decision boundary.* If a set of hyperplanes can separate between m given classes in R^n, these classes are *linearly separable.* Quite often a set of classes cannot be discriminated by linear decision functions. In this case we can either use *generalized decision functions* (nonlinear) in the original *pattern space*, i.e. use a *nonlinear classifier* or transform the problem to a space of a much higher dimension where classification is carried using linear boundaries.

Minimum-Distance Classifiers

If the training patterns seem to form clusters we often use classifiers which use distance functions for classification. If each class is represented by a single prototype called the *cluster center*, we can use a *minimum-distance* *classifier* to classify a new pattern. A similar modified classifier is used if every class consists of several clusters. The *nearest-neighbor* classifier classifies a new pattern by measuring its distances from the training patterns and choosing the class to which the nearest neighbor belongs.

Sometimes the *a priori* information is the exact or approximate number of classes c. Each training pattern is in one of these classes but its specific classification is not known. In this case we use algorithms to determine the cluster (class) centers by minimizing some performance index. These centers are found iteratively and then a new pattern is classified using a minimum-distance classifier. One such algorithm is c-Means where the exact number of classes is known. A more ambiguous situation is assumed by the ISODATA algorithm. We only have a *desired* number k of clusters and the final number of classes which is determined by the algorithm cannot be much higher or much lower than k.

Statistical Approach

Many times the training patterns of various classes overlap for example when they are originated by some statistical distributions. In this case a statistical approach is appropriate, particularly when the various distribution functions of the classes are known. A statistical classifier must also evaluate the *risk* associated with every classification which measures the probability of *misclassification*. The *Bayes classifier* based on Bayes formula from probability theory minimizes the total expected *risk*. To use Bayes classifier one must know a priori the pattern distribution function for each class. If these distributions are not known they must be approximated using the training patterns. Sometimes the functional form of these distributions is known and one must only estimate its parameters. However, in some applications even the distribution's form is unknown and must (approximately) be found. To do so we may for example perform functional approximation using expansions by orthogonal functions.

Fuzzy Classifiers

Quite often classification is performed with some degree of uncertainty. Either the classification outcome itself may be in doubt, or the classified pattern x may belong in some degree to more than one class. Fore example a person 5'8" tall does not fully belong to the class 'tall', yet at the same time he cannot be fully accepted in the class 'short' (provided that only these two classes exist). We thus naturally introduce *fuzzy classification* where a pattern is a member of every class with some grade of membership between 0 and 1. In this text we are mainly interested in fuzzy classification using equivalence relations and in fuzzy clustering. The *crisp c-Means* algorithm is generalized and replaced by the *fuzzy c-Means* and after the cluster centers are determined, each incoming pattern is given a final set of *grades of membership* which determine the degrees of its classification in the various clusters.

Syntactic Approach

Unlike the previous approaches, the *syntactic pattern recognition* utilizes the *structure* of the patterns. Instead of carrying an analysis based strictly on quantitative characteristics of the pattern, we emphasize the interrelationships between the *primitives*, the components which compare the pattern. Typical patterns which are subject to syntactic pattern recognition research are therefore characters, fingerprints, chromosomes, etc. The analogy between the structure of some patterns and the syntax of a language which has a solid theoretical basis is very attractive. By introducing the concept of a formal grammar and language we are able to design syntax classifiers that can classify a given pattern which is now presented as a string of symbols. In general, given a specific class, a grammar whose language consists of patterns in this class is designed. For an unknown new pattern a syntax classifier analyzes the pattern (a string) in a process called *parsing* and determines whether or not that string belongs to the language (class).

Neural Nets

The neural net approach assumes as other apporaches before that a set of training patterns and their correct classifications is given. The *architecture* of the net which includes *input layer*, *output layer* and *hidden layers* may be very complex. It is characterized by a set of weights and activation function which determine how any information (input signal) is being transmitted to the output layer. The neural net is trained by training patterns and adjusts the weights until the correct classifications are obtained. It is then used to classify arbitrary unknown patterns. There are several popular neural net classifiers, from the simple *perceptron* to the more advanced *backpropagation classifier.*

Pattern recognition and classification have been used for numerous applications. A detailed list is given below:

1. *Scientific Applications*:
   (a) Astronomy: telescope resolution improvements and atmospheric degradation removal.
   (b) Geology—planetary exploration: crater counts, color analysis, robotics, topography, atmospheric measurements and analysis, landing site and related evaluations, and terrestrial geologic feature analysis and charting.
   (c) Geology—cartography and geodesy: mosaicing, surface model fitting, and maps (making and alteration).
   (d) Bubble chamber tracking and electron microscope crystallography.
   (e) Satellite data analysis.
   (f) Sensing for life and date analysis on remote planets.
2. *Life and Behavioral Sciences*:
   (a) Anthropology.
   (b) Archeology.
   (c) Entomology.
   (d) Biology and botany: microbiology, ecology, and zoology.
   (e) Psychology: sociological aspects and criminological aspects.
   (f) Cybernetics.
   (g) Information management systems.
   (h) Education.
   (i) Communication.
3. *Industrial Applications*:
   (a) Character recognition.
   (b) Image controlled machines (process control).
   (c) Signature analysis.
   (d) Speech analysis.

   (e) Photographic recognition.
   (f) Mineral exploration (subsurface analysis).
   (g) Internal flow detection (X-ray and sonic).
   (h) Commercial photograph enhancement.
   (i) Multimedia and animation.
   (j) Electronic toys design.
   (k) Automated cytology.

4. *Medical Applications*:
   (a) Microscopic examination and biomedical data: blood cell counting and blood tests, cancer cell identification and tests, neuron measurements, chromosome karyotyping, bone composition analysis, automated focusing and positioning, and brain-tissue studies.
   (b) Radioisotope examination.
   (c) X-ray examination and tomography: blood vessel thickness measurements, heart size measurements, breast cancer detection, intracranial blood vessel constriction detection, dental charting and analysis, bone structure analysis, pulmonary disease diagnosis, and skeletal structure analysis.
   (d) Electrocardiogram and vectorcardiogram analysis.
   (e) Electroencephalogram tracing and neurobiological signal processing.
   (f) Drug interaction.
   (g) Chromosome properties for genetic studies.

5. *Agricultural Applications*:
   (a) Crop analysis.
   (b) Soil evaluation.
   (c) Process control.
   (d) Earth-resource photography.

6. *Governmental Applications*:
   (a) Weather prediction: cloud tracking and water temperature measurements.
   (b) Public systems: traffic analysis and control, urban growth determination, smog detection and measurement, and air traffic radar data reduction.
   (c) Earth-resource data and remote sensing.

7. *Some Specific Military Applications*:
   (a) Aerial photography and remote sensing.
   (b) Sonor detection and classification.
   (c) ATR: Automatic Target Recognition.

1.3 DATA MINING AND KNOWLEDGE DISCOVERY

Even though this text represents only the fundamental entities in the field of pattern recognition, we feel that it will not be complete without devoting a small section to the subject of data mining and knowledge discovery, in which classification plays a major role.

Throughout the text when we talk about classification, what we have in mind is the process assigning an item to its "natural group". In a more concrete sense, the objective of clustering is to sort a data set into categories such that the degree of "natural association" is high among members of the same category and low between members of different categories. In many cases, however, classification means *finding* the categories themselves from a given set of unclassified data. In essence this is what knowledge discovery and data mining is all about. When acquiring knowledge from data, the problem at times may be in the data itself, which may have limited breadth or coverage. While the development of databases has provided us with an effective tool for storage and lookup of large data sets, the issues related to knowledge discovery in these data glut, depends heavily on the field of pattern classification, since the notion of finding useful patterns (which in essence are just nuggets of knowledge) from raw data is the essence of information harvesting, which this text is all about.

Knowledge discovery (KD) and data mining (DM) systems draw upon methods and techniques from the field of pattern recognition, as well as related topics in database systems, artificial intelligence, machine learning, statistics, and expert systems, where the unifying goal is extracting knowledge from large volumes of data.

In their edited volume "Advances in Knowledge Discovery and Data Mining", Fayyad et.al. provide us with the following statement:

> *Knowledge discovery in databases (KDD) is the non-binial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data.*

They use the notion of *interestingness* to denote the *overall* measure of pattern value, combining validity, novelty, usefulness, and simplicity. The data mining step in the knowledge discovery process is therefore concerned with *means* by which patterns are extracted (as well as enumerated) from the raw data. In essence, the knowledge discovery process itself involves the evaluation and interpretation of the different patterns in order to provide decision-making with additional information on what constitutes knowledge and what does not. We should keep in mind, however, that in the context of knowledge discovery, *description* (finding human-interpretable patterns describing the data) tends to be more important than *prediction* (using some variables in the database in order to predict unknown other variables of interest), which is in contrast to pattern recognition, where prediction is usually the major goal of the analysis process.

One particular approach which we would like to mention is that of extracting fuzzy rules from raw data, which allows relationships in the raw data to be modelled by Fuzzy IF-THEN rules that are easy to validate and understand. Because fuzzy logic allows us to express nonlinear relationships by simple sets of qualitative IF-THEN rules, we can easily capture the essence of data behavior. Capturing that behavior, which is in essence knowledge discovery, in the form of Fuzzy IF-THEN rules, rather than by neural networks or surface approximation, provide us with a set of fuzzy rules which are easy to verify, validate, understand, explain and extend. This is a powerful framework not only for capturing the behavior of high dimensional data sets but also for explaining the behavior of the data sets, especially in non-stationary cases as well as in those cases where missing and noisy data is an acute pattern or when complex relationships exist between fields representing the data in a database. With the increase awareness of the advantages of representing classifiers in the form of sets of fuzzy IF-THEN rules, extracting fuzzy rules from raw data, with or without neural networks (for adaptive learning) will result in efficient and robust algorithms, especially for high dimensional and noisy data.

1.4 REFERENCES

For additional information on pattern recognition we refer the reader to the following sources. These include many classical monographs which explore in depth a variety of topics covered in our text.

[Agarwala 1977] AGARWALA, A.K., (ed.), *Machine Recogniton of Patterns*, IEEE Press, 1977.

[Anderberg 1973] ANDERBERG, M.R., *Cluster Analysis for Applications*, Academic Press, New York, 1973.

[Anderson/Rosefeld 1988] ANDERSON, J.A., AND E. ROSENFELD (eds.), *Neuro-computing: Foundations of Research*, MIT Press, Cambridge, Mass., 1988.

[Backer 1995] BACKER, E., *Computer-assisted Reasoning in Cluster Analysis*, Prentice Hall, Great Britain, 1995.

[Bezdek 1981] BEZDEK, J., *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.

[Bezdek/Pal 1992] BEZDEK, J.C. AND S.K. PAL (eds.), *Fuzzy Models for Pattern Recognition: Methods that Search for Pattern in Data*, IEEE Press, New York, 1992.

[Bow 1984] BOW, S.T., *Pattern Recognition*, Marcel Dekker, New York, 1984.

[Bunke 1982] BUNKE, H., 'Attributed Programmed Graph Grammars and Their Application to Schematic Diagram Interpretation,' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-4, No. 6, Nov. 1982, pp. 574-582.

[Carpenter 1989] CARPENTER, G., 'Neural Network Models for Pattern Recognition and Associative Memory,' *Neural Networks*, Vol. 2, 1989, pp. 243-257.

[Chen 1973] CHEN, C.H., *Statistical Pattern* Recognition, Hayden, Washington, D.D., 1973.

[Chien 1978] CHIEN, Y.T., *Interactive Pattern* Recognition, Marcel Dekker, New York, 1978.

[Coleman/Andrews 1979] COLEMAN, G.B., AND H.C. ANDREWS, 'Image Segmentation by Clustering,' *Proceedings of the IEEE*, Vol. 67, May 1979, pp. 773-785.

[Devijver/Kittler 1982] DEVIJVER, P., AND J. KITTLER, *Pattern Recognition: A Statistical Approach*, Prentice-Hall, Englewood Cliffs, N.J., 1982.

[Dubois/Pvade 1979] DUBOIS, D. AND H. PVADE, *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York, 1979.

[Duda/Hart 1973] DUDA, R.O., AND P.E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, New York, 1973.

[Everitt 1977] EVERITT, B., *Cluster Analysis*, Heinemann Educational Books, 1977.

[Fausett 1994] FAUSETT, L., *Fundamentals of Neural Networks*, Prentice Hall, Englewood Cliffs, N.J., 1994.

[Fayyad/et.al. 1996] FAYYAD, U.M. et. al. (eds.) Advances in Knowledge Discovery and Data Mining, AAAI Press/the MIT Press, Merile lPark, California, 1996.

[Fu 1968] FU, K.S., *Sequential Methods in Pattern Recognition and Machine Learing*, Academic Press, New York, 1968.

[Fu 1974] FU, K.S., *Syntactic Methods in Pattern Recognition*, Academic Press, New York, 1974.

[Fu 1980] FU, K.S., '*Recent Developments in Pattern Recognition*,' IEEE Transactions on Computers, Vol. C-29, No. 10, Oct. 1980, pp. 845-857.

[FU 1 1982] FU, K.S., *Syntactic Pattern Recognition and Applications*, Prentice-Hall, Englewood Cliffs, N.J., 1982.

[Fu 2 1982] FU, K.S. (ed), *Application of Pattern Recognition*, CRC Press, Cleveland, OH, 1982.

[Fu/Young 1985] FU, K.S., AND T.Y. YOUNG (eds.), *Handbook of Pattern Recognition and Image Processing*, Academic Press, New York, 1985.

[Fukunaga 1972] FUKUNAGA, K., *Introduction to Statistical Pattern Recognition*, Academic Press, New York, 1972.

[Gonzales/Thomason 1978] GONZALEZ, R.C., AND M.G. Thomason, *Syntactic Pattern Recognition*, Addison-Wesley, Reading, Mass., 1978.

[Haralick 1978] HARALICK, R.M., 'Structural Pattern Recognition, Homomorphisms, and Arrangements,' *Pattern Recognition*, Vol. 10, 1978, pp. 223-236.

[Hartigan 1975] HARTIGAN, J.A., *Clustering Algorithms*, John Wiley & Sons, New York, 1975.

[Hopfield/Tank 1986] HOPFIELD, J.J., AND D.W. TANK, 'computing with Neural Circuits: A Model,' *Science*, Vol. 233, Aug. 1986, pp. 625-633.

[Hwang et al. 1986] HWANG, V.S., L.S. DAVIS, AND T. MATUSUYAMA, 'Hypothesis Integration in Image Understanding Systems,' *Computer Vision, Graphics and Image Processing*, Vol. 36, 1986, pp. 321-371.

[Jain/Dubes 1988] JAIN, A.K., AND R. Dubes, Algorithms for Clustering Data, Prentice-Hall, Englewood Cliffs, N.J., 1988.

[Jang/et.al. 1997] JANG, J.-S.R., C.-T. SUN, AND E. MIZUTANI *Neuro-Fuzzy and Soft Computing*, Prentice Hall, Upper Saddle River, N.J., 1997.

[Kanal 1979] KANAL, L.N., 'Problem-solving Models and Search Strategies for Pattern Recognition,' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-1, April 1979, pp. 194-201.

[Kandel/Yelowitz 1974] KANDEL, A. AND Yelovitz, Fuzzy Chairs, IEEE Transactions on System, Man and Cybernetics, SMC-4, pp. 472-475, 1974.

[Kandel 1975] KANDEL, A., Fuzzy Hierarchical Classifications Of Dynamic Patterns, *Invited presentation at NATO A.S.I. on Pattern Recognition and Classification*, France, September 3-17, 1975.

[Kandel 1979] KANDEL, A., 'Fuzzy Techniques In the Clustering of Static And Dynamic Patterns', *Proceedings of the 1979 International Conference on Cybernetics and Society*, Denver, October 1979.

[Kandel 1982] KANDEL, A., *Fuzzy Techniques in Pattern Recognition*, Wiley Intersciences, New York, 1982.

[Kandel 1986] KANDEL, A., *Mathematical Techniques with Applications*, Addison-Wesley Publishing Co., Reading, Mass., 1986.

[Kandel et.al 1997] KANDEL, A., Y.Q. ZHANG, H. BUNKE, 'A Genetic Fuzzy Neural Network for Pattern Recognition,' *Proceedings of FUZZ-IEEE 1997*, pp. 11-21, July 1997, Barcelona, Spain.

[Khanna 1990] KHANNA, T., *Foundations of Neural Networks*, Addison-Wesley, Reading, Mass., 1990.

[Klir/Yuan 1995] KLIR, G.J. AND B. YUAN, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Upper Saddle River, N.J., 1995.

[Kohonen 1984] KOHONEN, T., *Self-Organization and Associative Memory*, Springer-Verlag, Berlin, 1984.

[Kosko 1990] KOSKO, B., 'Unsupervised Learning in Noise,' *IEEE Transactions on Neural Networks*, Vol. 1, No. 1, March 1990, pp. 44-57.

[Kosko 1991] KOSKO, B., *Neural Networks and Fuzzy Systems*, Prentice Hall, Englewood Cliffs, NY, 1991.

[Miclet 1986] MICLET, L., *Structural Methods in Pattern Recognition*, Springer-Verlag, New York, 1986.

[Pal/Majumder 1986] PAL, S.K., AND D.K.D. MAJUMDER, *Fuzzy Mathematical Approach to Pattern Recognition*, John Wiley & Sons, New York, 1986.

[Pal/Pal 1986] PAL, N.R., K. PAL AND J.C. BEZDEK, 'A Mixed *c*-Means Clustering Model,' *Proceedings of FUZZ-IEEE 1997*, pp. 11-21, July 1997, Barcelona, Spain.

[Pao 1989] PAO, Y.H., Adaptive *Pattern Recognition and Neural Networks*, Addison-Wesley, Reading, Mass., 1989.

[Patrick 1972] PATRICK, E. A., *Fundamentals of Pattern Recognition*, Prentice-Hall, Englewood Cliffs, N.J., 1972.

[Pavlidis 1977] PAVLIDIS, T., *Structural Pattern Recognition*, Springer-Verlag, New York, 1977.

[Pedrycz 1990] PEDRYCZ, W., *Fuzzy Sets in Pattern Recognition Methodology and Methods, Pattern Recognition*, vol. 23, pp. 121-146, 1990.

[Pedrycz 1995] PEDRYCZ, W., *Fuzzy Sets Engineering*, CRC Press, Boca Raton, Florida, 1995.

[Rosenfeld 1979] ROSENFELD, A., *Picture Languages*, Academic Press, new York, 1979.

[Rosenfeld 1984] ROSENFELD, A., *The Fuzzy Geometry of Image Subjects, Pattern Recognition Letters*, vol. 2, pp. 311-317, 1984.

[Ross 1995] ROSS, T.J., *Fuzzy Logic With Engineering Applications*, McGraw-Hill, Inc., 1995.

[Rummelhart/McClelland 1 1986] RUMMELHART, D.E., AND J.L. McCLELLAND, *Parallel Distributed Processing—Explorations in the Microstructure of Cognition, Volume 1: Foundations*, MIT Press, Cambridge, Mass., 1986.

[Rummelhart/McClelland 2 1986] RUMMELHART, D.E., AND J.L. McCLELLAND, *Parallel Distributed Processing—Explorations in the Microstructure of Cognition, Volume 2: Psychological and Biological Models*, MIT Press, Cambridge, Mass., 1986.

[Ruspini 1969] RUSPINI, E., 'A New Approach to Clustering,' *Inf. Control*, Vol. 15, pp. 22-32.

[Schalkoff 1989] SCHALKOFF, R.J., *Digital Image Processing and Computer Vision*, John Wiley & Sons, New York, 1989.

[Schalkoff 1992] SCHALKOFF, R.J., *Pattern Recognition: Statistical, Structural and Neural Approaches*, John Wiley & Sons, 1992.

[Simpson 1990] SIMPSON, P.K., *Artificial Neural Systems*, Pergamon Press, Elmsford, New York, 1990.

[Sklansky 1973] SKLANSKY, J. (ed.), *Pattern Recognition: Introduction and Foundations*, Dowden, Hutchinson and Ross, Straudsburg, Pa., 1973.

[Tamura/et al 1971] TAMURA, S., S. HIGUCHI, and K. TANAKA, 'Pattern classification Based on Fuzzy Relations,' *IEEE Trans. Syst. Man Cybern.*, vol. 1, pp. 61-66.

[Therrien 1989] THERRIEN, C.W., *Decision Estimation and Classification: An Introduction to Pattern Recognition and Related Topics*, John Wiley & Sons, New York, 1989.

[Thomason 1975] THOMASON, M.G., *Finite Fuzzy Automata, Regular Languages and Pattern Recognition*, Pattern Recognition, vol. 5, pp. 383-390, 1975.

[Tou/Gonzalez 1974] TOU, J., AND R. GONZALEZ, *Pattern Recognition Principles*, Addison Wesley, Reading, Mass., 1974.

[Turksen 1993] Turksen, I.B. AND S. Jiang, *Rule Base Reorganization and Search with a Fuzzy Cluster Analysis*, International Journal of Approximate Reasoning, vol. 9., no. 3, pp. 1267-196, 1993.

[Watanable 1985] WATANABE, S., *Pattern Recognition: Human and Mechanical*, John Wiley & Sons, New York, 1985.

[Yager/Zadeh 1992] YAGER, R.R. AND L.A. Zadeh (eds.), An Introduction to Fuzzy Logic Applications in Intelligent Systems, Kluwer, Boston, 1992.

[Yager/Zadeh 1994] YAGER, R.R. AND L.A. ZADEH (eds.), *Fuzzy Sets, Neural Networks and Soft Computing*, Von Nostrand Reinhold, New York, 1994.

[Yager/Filev 1994] YAGER, R.R. AND D.P. FILEV, *Essentials of Fuzzy Modeling and Control*, John Wiley & Sons, New York, 1994.

[Young/Calvert 1974] YOUNG, T.Y., AND T.W. CALVERT, *Classification, Estimation and Pattern Recognition*, Elsevier, New York, 1974.

[Zadeh 1965] ZADEH, L., 'Fuzzy Sets,' *Inf. Control*, 8 pp. 338-353, 1965.

[Zadeh 1971] ZADEH, L., 'Similarity Relations and Fuzzy Orderings,' *Inf. Sci.*, vol. 3, pp. 177-200.

[Zadeh 1975] ZADEH, L., et.al., *Fuzzy Sets and Their Applications to Cognitive and Decision Processes*, Academic Press, New York, 1975.

[Zimmerman 1985] ZIMMERMAN, H.J., *Fuzzy Set Theory and its Applications*, Kluwer, Boston, 1985.

[Zimmerman 1987] ZIMMERMAN, H.J., *Fuzzy Sets, Decision Making and Expert Systems*, Kluwer, Boston, 1987.

The following journals contain major papers on the theory of pattern recognition and its applications:

Pattern Recognition
Pattern Recognition - letters
IEEE Transactions on Pattern Analysis and Machine Intelligence
IEEE Transactions on Neural Networks
IEEE Transactions on Fuzzy Systems
IEEE Transactions on Systems, Man and Cybernetics
Fuzzy Sets and Systems
Information Sciences
International Journal of Pattern Recognition and Artificial Intelligence