PREFACE

Microarray technology allows one to measure gene expression levels simultaneously on the whole genome scale. The rapid progress generates a great wealth of information and challenges in making inferences from such massive data sets. The greatest challenges in microarray technology are that the number of variables far exceeding the number of samples, Standard statistical methodologies may not work well or may give poor results. Modifications of existing methodologies or development of new methodologies is needed for the analysis of microarray data.

Focus of this research work is to collect and unify robust statistical methods for microarray data analysis in identification of differentially expressed genes. The major attention is to propose new efficient statistical methods for double channel microarray data and comparing the performance with other available methods. The thesis is organized into 6 chapters as follows.

Chapter 1 deals with introduction, era of bioinformatics and microarray technology, biological background, micro array experimental design. And it also briefly explains the statistical issues and computational complexity of microarray data.

Chapter 2 presents literature review of existing statistical approaches in microarray data analysis and the frequently used visualization tools for gene expression data. And also draws attention to preprocessing and normalization techniques for microarray data to remove systematic sources of biases.

Chapter 3 discusses unsupervised statistical methods to identify differentially expressed genes. This chapter explains dimension reduction methods and clustering methods, and how these methods are applied to double channel microarray data sets.

Chapters 4, discuss different supervised learning methods namely RBF, MLP and SVM for the high dimensional data analysis. And compared the performance of different supervised learning methods in the classification study of the complex two channel gene expression data sets. A fuzzy based model is also proposed for the classification of throughput data sets.
Chapter 5 focuses on the recent advances in statistical methods for microarray data analysis based on Hidden Markov Model and MCMC based Bayesian methods for exploring high dimensional expression data sets.

Chapter 6 presents discussion, conclusion and topics for future research.