Using this last approach to cancer tumor classification, Mitra et al. (2008) provide practical examples of and benchmark comparisons to earlier less-accurate methods of classification (see pages 316+), to such cancers as leukemia and colon tumors.

POSTSCRIPT

Statistical and data mining analysis in bioinformatics is still in somewhat of a chaotic state of trial and error. Researchers in this field are still debating the value of this or that method; the field does not yet have a standard methodology of analyzing its very large and complex databases. As Parida (2007) pointed out in the introductory sentences of her book:

Major scientific discoveries have been made quite by accident; however, a closer look reveals that the scientists were intrigued by a specific PATTERN in the observations; an excellent example is Edward Jenner and the discovery of how milkmaids in England developed immunity to smallpox, followed by the development of the smallpox vaccination, and the eventual eradication, a few years ago, of smallpox on planet Earth. (page 1; paraphrase)

Pattern discovery is the operation that data mining and text mining excel in doing. These tools are necessary in the study of any complex bioinformatics phenomena. Ideally, in the future, a data mining/pattern recognition system will have standardized components, bringing bioinformatics and DNA microarray analysis into a more mature field, and providing practical ways of making accurate decisions in many areas, including disease diagnosis with accompanying action plans with a known probability of success.

Tutorial Associated with This Chapter on Bioinformatics

On the accompanying DVD packaged with this book, go to the Tutorial section and select the “Tutorial—CancerGene,” where a K-nearest neighbors Bayesian model is used to show a 99% accuracy prediction of breast cancer.

A “Text Mining in Bioinformatics & Medical Informatics” white paper is included on the DVD; go to the TUTORIAL – PDF_PPT_ETC folder to find this paper; it contains a good discussion of the issues surrounding text mining in the bioinformatics field.

Books, Associations, and Journals on Bioinformatics, and Other Resources, Including Online

Associations and Organizations

- EMBnet (http://www.embnet.org/)
- European Bioinformatics Institute (http://www.ebi.ac.uk/)
- European Molecular Biology Laboratory (http://www.ebi.ac.uk/)
- The International Society for Computational Biology (http://www.iscb.org/)