INTERNATIONAL CENTRE
OF EXCELLENCE FOR
EDUCATION IN
MATHEMATICS

Non-linear Optimisation in Classification and Gene Selection

Michelle Dunbar, School of Mathematics and Statistics, University of New South Wales

The recent development of biological techniques such as DNA microarray data technology has allowed the study and analysis of the expression levels of potentially thousands of genes *simultaneously*. While this provides us with a great deal of useful information, conventional methods of analysis are often limited, owing to the inherent problem of possessing very high dimensional–low sample size data.

One method proposed to deal with this problem is known as "gene selection", and involves selecting the "most relevant" genes for classification purposes. Knowledge of the "most relevant" genes may allow us to gain an insight into the biology of certain diseases (such as cancer) and may eventually aid its practicality as a clinical device.

Various non-linear optimisation models (successfully used in the classification of a number of diseases) have recently been adapted to study the problem of gene selection for classification. Although gene selection and classification are two closely related problems, most existing approaches handle them separately by selecting genes prior to, or immediately after classification.

The aim of my research project was to firstly gain an understanding of various non-linear classification methods, and secondly to investigate a number of fairly *recently proposed* methods (ie. 2004 - 2006) for *simultaneous* gene selection and classification that have shown promise in not only improving the accuracy of the resulting classifier, but a reduction in the time required to solve the problem. These methods involve the formulation of a non-linear optimisation problem, and are unfortunately very difficult to solve - As with the exception of a few well known classes of problems, there is no *general* procedure for solving a non-linear problem. Moreover, a particular method involved the formulation of a non-convex, non-differentiable function, adding to the complexity. In light of this information, I spent the remainder of the project reviewing various techniques and alternative formulations that may be used to overcome these problems.

Over the course of my honours year I hope to further analyse and build upon this work, as well as apply these methods to a real world dataset and compare the results with existing methods.

I found this project extremely interesting, and it has given me insight into how research is conducted by mathematicians. At the conclusion of my vacation scholarship, I had the opportunity to present my work at the "Big Day In". I found this two-day event very rewarding, not only having the opportunity to develop my skills in communicating mathematics, but to meet fellow students studying a wide range of extremely interesting topics!

I'd like to thank AMSI for this scholarship, the School of Mathematics and Statistics, UNSW and my supervisor A/Prof Jeyakumar for his guidance over the six weeks.

Michelle received an ICE-EM Vacation Scholarship in December 2006.
See http://www.ice-em.org.au/students.html#scholarships2007