

Department of Applied and Computational Mathematics and Statistics Colloquium

Omar De La Cruz Cabrera

Department of Epidemiology and Biostatistics
Case Western Reserve University

will give a lecture entitled:


Kernel and Geometric Methods in the Analysis of Genetic Data

Abstract

Kernels allow us to study non-linear features of data sets using classical linear statistical methods. In the setting of exploratory data analysis, in which the interpretation of the results of well-known classical methods like Principal Components Analysis can be tricky, the use of kernels can make things even more difficult.

I will discuss some difficulties that arise in the use of kernel methods for exploratory data analysis, especially in the interpretation of results, and show how some of those difficulties appear even in classical methods when the data is very high dimensional, as is the case of genotype data. We will then describe a statistical model that can help understand what kernel methods actually do, in general, as well as provide some guidance in the choice of method and of kernel, and help interpret the results of particular analyses.

In the second part of the talk I will describe the use of methods from manifold learning and computational topology in a particular example: the detection of genes whose expression is regulated by the cell division cycle, using data without explicit time information. Our approach makes it possible to find cycle-regulated genes in different tissues, using cells extracted from living individuals, without need for culturing and synchronization. This has the potential for elucidating differences in the mechanism of cell division and proliferation as it happens in, for example: embryos, adult epithelia,



**Wednesday, February 1, 2012
4:00 p.m. to 5:00 p.m.
127 Hayes-Healy Center**