C R Rao Advanced Institute of Mathematics, Statistics and Computer Science

University of Hyderabad Campus, Prof. C.R. Rao Road, Gachibowli, Hyderabad-500046, Phone-040-23013118

Seminar Notice

Predictive problems in high-dimensional data: Some examples

Friday, 10th February, 2012 at 4 PM at Aryabhtta Conference Hall,
C R Rao Advanced Institute of Mathematics, Statistics and Computer Science

by

Dr. Madhuchhanda Bhattacharjee

Reader, Department of Statistics, University of Pune, India

Abstract

An emerging and important area of research is to relate high-dimensional genetic or genomic data to various (clinical) phenotypes of patients. For dichotomous/categorical responses this is typically known as classification problem and has been looked into by experts of many different fields, with statistics playing only a minor role therein. Often the underlying response or phenotype is continuous and these are recoded into categories for ease of interpretation. However many situations require dealing with the actual phenotype in its continuous form and we review certain aspects of modelling and prediction of such response variables. Along with the use of whole repertoire of regression models, several applications of survival models have also been proposed. As is seen elsewhere as well, predominantly models are formed in classical statistical framework compared to Bayesian setup. We investigate applications of some these models along with new (Bayesian) models proposed by us with applications to real data. Apart from the regular issues that are discussed when the data has large number of predictors compared to number of samples, there are aspects that are easily overlooked pertaining to prediction in such situations. Since the objective is to develop an accurate predictor which can be adopted in practice (in real life) it becomes essential to identify proper method of validation (and needless to say carry out validation accordingly). Often mistakes occur, firstly because of lack of understanding of difference between model fitting and deriving a predictive model, secondly inadequate validation system. These render such a model either useless or disappointing in real life prediction. We conclude with some discussion on this topic including few of our observations on this matter.

***All are cordially invited ***

Director, AIMSCS