A LiNKAGE OF COX PROPORTIONAL HAZARD MODEL AND PARTIAL LEAST SQUARES TO ANALYZE GENE EXPRESSION DATA WITH PATIENT SURVIVAL TIME

Jean Kongpinda & Venessa Tavares

Abstract

Motivation: Microarrays have been increasingly used in biomedical research as a way of finding the relationship between the genotype and phenotype of an individual. Doctors would like to predict the physical characteristics of a patient (phenotype) after analyzing the expression of his genes (genotype). This has many applications in medicine and has been used on many cancer patients. This method would help to determine whether a patient has cancer or not or what stage of cancer he belongs to. Most researches have been done on those types of phenotype, but less has been done in the case where the phenotype of the patients in study is their survival time. A simple regression model (such as a linear regression model) can be applied on the former one, but will not be appropriate with the latter one: this is mainly due to the problem of censoring. In this paper we consider the case where the phenotype is the survival time. We examine the problem of survival analysis when the number of variables is much larger than the sample size, a typical case in DNA microarray data.