



Clustering longitudinal profiles using P-splines and mixed effects models applied to time-course gene expression data

John Hinde*

NUI Galway, Galway, Ireland – john.hinde@nuigalway.ie

Norma Bargary

University of Limerick, Limerick, Ireland – Norma.Bargary@ul.ie

Emma Holian

NUI Galway, Galway, Ireland – emma.holian@nuigalway.ie

Longitudinal data is becoming increasingly common and various methods have been developed to analyze this type of data. Profiles from time-course gene expression studies, where cluster analysis plays an important role to identify groups of co-expressed genes over time, are investigated. A number of procedures have been used to cluster time-course gene expression data, however there are many limitations to the techniques previously described. An alternative approach is proposed, which aims to alleviate some of these limitations. The method exploits the connection between the linear mixed effects model and P-spline smoothing to simultaneously smooth the gene expression data to remove any measurement error/noise and cluster the expression profiles using finite mixtures of mixed effects models. This approach has a number of advantages, including decreased computation time and ease of implementation in standard software packages.

Keywords: longitudinal data; smoothing; finite mixture models; classification.