

Yale University

From the Selected Works of Shuangge Ma

2012

Integrative analysis of multiple cancer genomic datasets under the heterogeneity model

Shuangge Ma, *Yale University*



Available at: <https://works.bepress.com/shuangge/44/>

Integrative analysis of multiple cancer genomic datasets under the heterogeneity model

Jin Liu¹, Jian Huang² and Shuangge Ma¹

¹Department of Biostatistics, School of Public Health, Yale University, 60 College Street, New Haven, CT 06520

²Department of Statistics & Actuarial Science, Department of Biostatistics, University of Iowa, Iowa City, IA 52242

Abstract. In the analysis of cancer studies with high-dimensional genomic measurements, integrative analysis provides an effective way of pooling information across multiple heterogeneous datasets. The genomic basis of multiple independent datasets, which can be characterized by the sets of genomic markers, can be described using the homogeneity model or heterogeneity model. Under the homogeneity model, all datasets share the same set of markers associated with responses. In contrast, under the heterogeneity model, different studies have overlapping but possibly different sets of markers. The heterogeneity model contains the homogeneity model as a special case and can be much more flexible. Marker selection under the heterogeneity model calls for bi-level selection to determine whether a covariate is associated with response in any study at all as well as in which studies it is associated with responses. In this study, we consider two minimax concave penalty (MCP) based penalization approaches for marker selection under the heterogeneity model. For each approach, we describe its rationale and an effective computational algorithm. We conduct simulation to investigate their performance and compare with the existing alternatives. We also apply the proposed approaches to the analysis of gene expression data on multiple cancers.

Keywords: Integrative analysis; Heterogeneity model; Marker selection.