
SHOFIQL ISLAM, McMaster University

Comparing the Performance of Linear and Nonlinear Principal Components in the Context of High Dimensional Genomic Data Integration

We integrate phenotype, gene expression and miRNA expression data to classify tumor or death of Lung Cancer patients. We apply linear and kernel principal component analysis (PCA) to reduce the dimension and subsequently integrate three domains based on several nonlinear approaches including logistic regression. We also conducted an extensive simulation to compare the performances. We observe that a reduced set of kernel principal component carries more information from a genetic process but failed to show better predictive ability in this occasion. Reducing dimension with linear PCA and a logistic regression model for classification seems to be sufficient for this purpose.