



MICROARRAY DATA: DIMENSION REDUCTION STRATEGIES

John Ratanaprasatporn *Binghamton University*
Israel Cabello *Texas State*

Abstract of Poster: This project is an extension of a collaborative research effort comparing the effectiveness of two dimension reduction strategies when applied to microarray data. We compare two methods: principal component analysis (PCA) and partial least squares (PLS). We hypothesized that (1) PLS would more efficiently model survival time, and (2) the genes highly weighted by PLS would be more relevant to survival time. We used the Cox Proportional Hazard Model to find the components that are significant predictors of survival time in each reduced dataset. We found more significant p-values for PLS and found that PLS could be accurately modeled using fewer components, which confirmed part (1) of our hypothesis. This poster looks more closely at the second part of our hypothesis. We wished to examine the genes selected by each process. We examined the twenty genes most heavily weighted by each method and looked at the ratios of the weights for each top gene to see whether the overlapping genes were weighted similarly. Overall, our results indicate that PCA is able to select some of the same significant genes identified by PLS, but PLS provides a much more concise representation of the data.