Modeling Pharmacogenomics of the NCI-60 Anticancer Data Set: Utilizing kernel PLS to correlate the Microarray Data to Therapeutic Responses

Nilanjan Dasgupta[1], Simon M. Lin[2], Lawrence Carin[1]

[1]EE Deptt. Duke University; [2]Duke Bioinformatics Share Resource, Duke University Medical Center

Modeling the relationship between the genomic features and the therapeutic responses is of central interest in pharmacogenomics [1]. The NCI-60 data set with both gene expression and drug activity measurements provides an excellent opportunity for this modeling exercise. The goals of this work are

1. To identify the genetic components contributing to drug sensitivity: in particular, we are going to identify the single drug resistance genes and multiple drug resistance genes.
2. To model the quantitative relationship between the microarray profile and the drug response.

To reduce noise in the data set, and to remove high multicollinearity among the gene expressions, we first reduce the dimensionality of the gene expression pattern using a kernel based PCA. To correlate the gene expression profile with the drug activity pattern, we utilize a soft modeling technique called Partial Least Squares (PLS) [2]. Soft modeling requires less stringent assumptions about the data than other modeling techniques [5]. In both PCA and PLS, we used kernel-based methods to exploit the non-linear dependence between the input and output data set. This is facilitated by a kernel function that implicitly carries out the regression in a higher-dimensional space where the data is linear [3].

The potential application of our model could be:

1. To predict the drug response of a new cell line. This result can be used for the choice of drug treatment.
2. To make an informed decision on the combination therapy to maximize the cytotoxic effect and minimize the drug-resistance potential.

The insights from this modeling study are going to be validated in the Cancer Center laboratories at the Duke University Medical Center. This is a work in progress. We hope to show the results at the CAMDA conference.