Network Topology Identification based on Measured Data

Mahdi Zamanighomi

We consider modeling of systems and learning of models from a limited number of measurements. As an inspiring example, a growing interest in biology is to determine dependencies among genes. Such problem, known as gene regulatory network inference, often leads to identifying of large networks through relatively small gene expression data.

The main purpose of the thesis is to develop generic models and learning methods for data based applications. In particular, we first build a dynamical model for gene-gene interactions to learn the topology of gene regulatory networks from gene expression data. Our proposed model is applicable to such complex gene regulatory networks that contain loops and non-linear dependencies between genes. We seek to use dynamical gene expression data when a system is perturbed. Ideally, such dynamical changes result from local genetic or chemical perturbations of systems in steady state that can be captured in a time-dependent manner. We present a low-complexity inference method that can be adapted to incorporate other information measured across a biological system. The performance of our method is examined employing both simulated and real datasets. This work can potentially inform biological discovery relating to interactions of genes in disease-relevant networks, synthetic networks, and networks immediate to drug response.

We next seek to estimate high-dimensional covariance matrices based on a few partial observations. Notably, covariance matrices can be utilized to form networks or improve network inference. We assume that the true covariance matrix can be modeled as a sum of Kronecker products of two lower dimensional matrices. To estimate covariance, we propose a convex optimization approach computationally affordable in high-dimensional setting and applicable to missing data. Regardless of whether the process producing missing values is random or not, our novel scheme can be used without employing any imputation methods. We characterize the symmetry and positive definiteness of the estimated covariance and further shed light on its square error performance. The effect of missing values on the estimation error is mathematically presented and numerical results are illustrated to validate our method.