DOWNLOAD PDF CONDITIONAL ASSUMPTION ON GRAPHICAL MODELS

Chapter 1 : Estimating mixed graphical models | R-bloggers

Probabilistic graphical models represent large joint distributions compactly using a set of \local" relationships speci ed by a graph. Each random variable in our model corresponds to a graph node.

Received Sep 18; Accepted Dec 1. The use, distribution or reproduction in other forums is permitted, provided the original author s or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms. This article has been cited by other articles in PMC. Abstract It is challenging to identify meaningful gene networks because biological interactions are often condition-specific and confounded with external factors. It is necessary to integrate multiple sources of genomic data to facilitate network inference. For example, one can jointly model expression datasets measured from multiple tissues with molecular marker data in so-called genetical genomic studies. In this paper, we propose a joint conditional Gaussian graphical model JCGGM that aims for modeling biological processes based on multiple sources of data. This approach is able to integrate multiple sources of information by adopting conditional models combined with joint sparsity regularization. We apply our approach to a real dataset measuring gene expression in four tissues kidney, liver, heart, and fat from recombinant inbred rats. Our approach reveals that the liver tissue has the highest level of tissue-specific gene regulations among genes involved in insulin responsive facilitative sugar transporter mediated glucose transport pathway, followed by heart and fat tissues, and this finding can only be attained from our JCGGM approach. Introduction Inference of gene networks plays an important role in revealing the interactions among genes that may lead to a better understanding of molecular mechanisms in organisms. Biologists routinely use high-throughput technologies e. Statisticians are often charged to explore interactions among genes through statistical analysis of these large data sets. It is natural to use multivariate approaches to analyze these high-throughput datasets, because multivariate methods may reveal various interactions among genes that cannot be captured from individual gene based approaches. In this paper we focus on a graphical model approach that aims at finding relationships among a group of genes, where a graph is used for encoding relationships among multiple variables. When a graph is used for a gene network, nodes represent genes and edges represent relationships between the connected genes. The edges can be defined with various relationships among genes. Therefore, when the expression profiles of two genes are correlated because they are both regulated by some other genes, the graphical model does not put an edge between these two genes because they are conditionally independent given the expressions of the common regulatory genes. In this way, the graphical model produces a more parsimonious graph than a relevance network. Gene network inference is a complex problem, because the relationships of genes are often affected by external variables e. This means that a single network inferred from gene expression measurements alone may not be adequate to describe the relationships among genes. Further, it is often desirable to jointly model gene networks under various conditions rather than considering them separately, because large parts of the networks are likely to share common topologies corresponding to similar underlying biological processes across conditions e. Therefore, one may want to infer multiple condition-specific networks in a single model framework, while the network models may also need to incorporate all available external variables as well. Such inference is possible through the analysis of datasets in genetical genomic studies from same genetic origin Jansen and Nap, where gene expressions from multiple tissues, as well as marker genotypes, are measured from the same set of individuals. These data allow us to perform an integrative analysis via joint conditional Gaussian graphical models JCGGM to infer relationships among genes. What types of biological networks have been inferred in the paper? We use gene expression data and marker data from recombinant inbred rats and infer gene regulation network by using genes consisting of the insulin responsive facilitative sugar transporter mediated glucose transport pathway. Our JCGGM found that the liver network has the highest tissue specificity, and this is in line with the role of SLC2A4 protein, which forms glucose concentration gradient of muscle and fat cells, as well as the specialized glycogen breakdown of glycogen phosphorylase that only occurs in liver tissue Watson et al. How were these networks validated? We

have performed simulation study to test performance of the proposed JCGGM approach and our approach performs the best over all simulation scenarios. We have also provided the scientific literature to support the validity of the inferred networks. In Section 3, we show the performance of our approach via a simulation study and then apply it to a genetical genomics study, where gene expressions from four different tissues are measured together with genotype data from recombinant inbred rats. The discussion follow in Section 4. Materials and methods 2. Material For a real data analysis, we used a dataset of Petretto et al. This strain was derived from a cross between the spontaneously hypertensive rat SHR and the brown norway BN strains Hubner et al. We downloaded the dataset normalized by the robust multi-array average RMA algorithm from www. From the same website, we also downloaded a genetic marker dataset that consists of markers. Methods In this section, we briefly introduce recent approaches for CGGMs as well as those for joint estimation of multiple Gaussian graphical models. We then propose a new method to combine these approaches in order for inferring networks from multiple sources of biological data for finding multiple CGGMs. Finally, we explain the simulation process for generating datasets that are used for comparing the performance of our proposed method. Hence, conditional independence can be directly inferred from zero entries of a precision matrix, when a multivariate Gaussian assumption is made. This model is called a GGM Lauritzen, Finding a sparse precision matrix with various regularizations such as lasso and adaptive lasso Tibshirani, ; Zou, has been studied by many researchers including Li and Gui ; Yuan and Lin ; Friedman et al. More recently, it has been noted that one can further elaborate a GGM by using extra sources of information. When the marker effect is ignored, there are two edges in the unconditional graphical model: After considering the marker effect, there is only one edge, represented by the solid line, in the conditional graphical model.

DOWNLOAD PDF CONDITIONAL ASSUMPTION ON GRAPHICAL MODELS

Chapter 2 : Probabilistic Graphical Models

An introduction to graphical models Graphical models are a marriage between probability theory and graph theory. depending on which conditional assumptions we.

The primary aim of this paper is to show how graphical models can be used as a mathematical language for integrating statistical and subject-matter information. In particular, the paper develops a principled, nonparametric framework for causal inference, in which diagrams are queried to determine if In particular, the paper develops a principled, nonparametric framework for causal inference, in which diagrams are queried to determine if the assumptions available are sufficient for identifying causal effects from nonexperimental data. If so the diagrams can be queried to produce mathematical expressions for causal effects in terms of observed distributions; otherwise, the diagrams can be queried to suggest additional observations or auxiliary experiments from which the desired inferences can be obtained. Causal inference, graph models, interventions treatment effect 1 Introduction The tools introduced in this paper are aimed at helping researchers communicate qualitative assumptions about cause-effect relationships, elucidate the ramifications of such assumptions, and derive causal inferences from a combination Inference in belief networks: Belief networks are popular tools for encoding uncertainty in expert systems. These networks rely on inference algorithms to compute beliefs in the context of observed evidence. In this document, we provide a self-contained, procedural guide to understanding and implementing PPTC. We synthesize various optimizations to PPTC that are scattered throughout the literature. We hope that this document makes probabilistic inference more accessible and a ordable to those without extensive prior exposure. Show Context Citation Context A graph-theoretic relation known as d-separation captures all such derivable independences encoded by the DAG [13]. This paper explores the role of Directed Acyclic Graphs DAGs as a representation of conditional independence relationships. We show that DAGs offer polynomially sound and complete inference mechanisms for inferring conditional independence relationships from a given causal set of such relationship We show that DAGs offer polynomially sound and complete inference mechanisms for inferring conditional independence relationships from a given causal set of such relationships. As a consequence, d-separation, a graphical criterion for identifying independencies in a DAG, is shown to uncover more valid independencies then any other criterion. In addition, we employ the Armstrong property of conditional independence to show that the dependence relationships displayed by a DAG are inherently consistent, i. Dependency knowledge of the form "x is independent of y once z is known" invariably obeys the four graphoid axioms, examples include probabilistic and database dependencies. Often, such knowledge can be represented efficiently with graphical structures such as undirected graphs and directe Often, such knowledge can be represented efficiently with graphical structures such as undirected graphs and directed acyclic graphs DAGs. In this paper we show that the graphical criterion called d-separation is a sound rule for reading independencies from any DAG based on a causal input list drawn from a graphoid. The rule may be extended to cover DAGs that represent functional dependencies as well as conditional dependencies.

Chapter 3 : Joint conditional Gaussian graphical models with multiple sources of genomic data

- Conditional independence (CI) assumptions useful - Recall NaÃ¯ve Bayes uses CI assumptions, but is extreme! - Graphical models express more flexible sets of conditional.

Chapter 4 : CiteSeerX â€” Citation Query Conditional independence in statistical theory

The graphical model represented by Gis then the set of distributions over Xthat satisfy the conditional independence assumptions specii¬•ed by the graph G. There has been a considerable line of work on learning parametric families of such graphical model.