

Assignment 3

Due: November 15th at 8pm EST

Instructions: This is an individual assignment, not group work. Though you may discuss the problems with your classmates, you must solve the problems and write the solutions independently. As stated in the syllabus, copying code from a classmate or the internet (even with minor changes) constitutes plagiarism. You are required to submit your answers in pdf form (use \LaTeX) in a file called `<your-UNI>-hw2.pdf` to courseworks. The code for the programming assignment should be appended at the end of this pdf. Late submissions will be penalized, except in extenuating circumstances such as medical or family emergency. Submissions submitted 0-24 hours late will be penalized 10%, 24-48 hours late by 20%, 48-72 hours late by 30%, and later than 72 hours by 100% (i.e., zero credit). Questions 1 and 2 are worth 7 points, question 3 is 16 points, for a total of 30 points.

Problem 1

Consider the simple 3-variable model in Figure 1 where all variables (including the latent L_1) are jointly Gaussian. Recall that this implies that the variables are related by linear structural equations, the coefficients of which are α, β, γ . Show that *any* value for the covariance (and hence correlation) between observed variables X_1 and X_2 can be achieved by some setting of the parameters β, γ . Explain the significance of this fact.

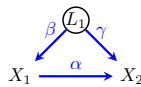


Figure 1

Problem 2

Consider the latent variable DAG in Figure 2. What would a CPDAG-learning algorithm (such as PC, GES, or similar) learn from data generated according to this structure, in the limit of infinite sample size? (That is, assume PC makes *no mistakes* about conditional independence relations among the *observed* variables. Obviously it only “knows” about observed variables, not anything latent.) Draw the implied CPDAG (i.e., what PC would learn in the limit), and summarize the output in words. How would you interpret this result? Is there something “right” or “wrong” about it (from an independence perspective and/or from a causal perspective)? What would FCI learn? You may ignore the propagation of the orientation rules (R1-R10) at the end of FCI – these are not important here. Finally, what would an algorithm that learns a Markov Random Field (undirected graph) learn from data generated according to this structure? Answer the same questions for the structure in Figure 3.

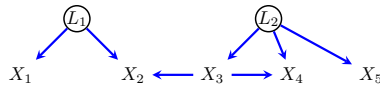


Figure 2

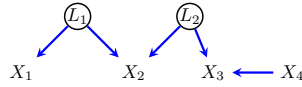


Figure 3

Problem 3

On Courseworks you'll find a .zip folder with some simulated data, generated from a Gaussian graphical model with 20 variables. Use the PC algorithm (in the R package `pcalg`) to estimate a CPDAG from this data. Use the glasso (graphical lasso) method to estimate an undirected graph from this data (use the R package `huge`)¹. Both algorithms depend on a tuning parameter: α in the PC algorithm corresponds to the conditional independence test threshold; λ in glasso corresponds to the sparsity penalty (effectively a threshold for deciding when the elements of the precision matrix are “small” enough to count as zero). Explore some ways of choosing the tuning parameter, and show how the estimated graph (or some features of the graph, like the number of missing edges) varies as you vary the tuning parameter. Explain every step of what you do with this data, visualize the results, and include your code.

¹The `huge` package implements a number of ways to estimate undirected graphs, including glasso. There is a document which explains the features of this package here: <https://CRAN.R-project.org/package=huge> (see “reference manual” and “vignette”)