NOTES. **NO** late submission will be accepted. Computer generated output without detailed explanations and remarks will not receive any credit. You may type out your answers, but make sure to use different fonts to distinguish your own words from computer output. Scan or take photos of your homework solutions and submit through CANVAS. For the simulation and data analysis problems, keep the code you develop as you may be asked to present your work in a in-class workshop.

1. This problem considers the three DAGs we have talked about in the lecture with more depth.
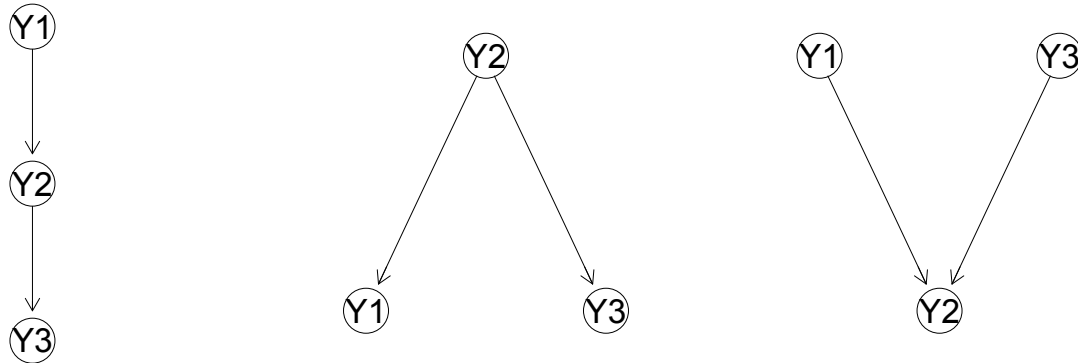
   (a) Write down the regression representation of the Gaussian graphical model for each of the given graph.

   (b) Write down the covariance matrix of the joint distribution corresponding to each of these three distributions.

   (c) Are any of these graphical models equivalent? [We say two models are equivalent, if any covariance matrix under one model can also be generated from the other model by selecting suitable parameters, and vice versa.] Prove your finding.

2. This problem uses a dataset `gene.txt` containing expression measurements of 9 genes. The names of the genes are included in the dataset. It also uses a R function `is.acyclic()`. Here is how you will use this function to tell whether a directed graph is acyclic or not. Represent a directed graph $\mathcal{G}$ as an adjacency matrix $\boldsymbol{A}$: if there is an edge from node i to node j, then $\boldsymbol{A}[i,j] = 1$, otherwise $\boldsymbol{A}[i,j] = 0$. Note that the adjacency matrix $\boldsymbol{A}$ is not symmetric for a directed graph. The function `is.acyclic()` takes the adjacency matrix $\boldsymbol{A}$ as the input and returns whether it is acyclic or not. Both the dataset and the R function can be downloaded from the Homework folder. The R file also contains some examples on how to use the `is.acyclic()` function.

   (a) Write you own code try all possible DAGs on the four genes `GAL1, GAL2, GAL3, GAL7`. There are 6 pairs of nodes, and between each pair of nodes, there can be three options: no edge, or an edge with either direction. Therefore, there are in total $3^6 = 729$ directed graphs over 4 nodes. For each of them, check whether the graph is acyclic using the `is.acyclic()` function. If it is indeed a DAG, estimated the model parameters, and report the BIC.

   (b) Report the models with three smallest BIC (there may be multiple models with the same BIC, report all of them). Plot the corresponding DAGs. [You can try to use the R code at the end of the slides to plot the DAGs. Or since there are only 4 nodes, you can draw by hand.]

3. This is a continuation of Problem 2.

   (a) Write you own code to implement the step-wise algorithm for estimating the graph structure. Write a generic code which takes a data matrix of any size as the input (so that it won't take any additional effort if you want to use it for the in-class workshop).

   (b) Try your program on the full `gene.txt` dataset. You will need an initial graph to start your algorithm, which can be the empty graph, full graph, or any graph you report in Part (b) of Problem 2. Report the DAG you found with the smallest BIC.

4. Let $\boldsymbol{x} = (x_1, \ldots, x_p)'$ be a random vector with multivariate normal distribution $N(\boldsymbol{\mu}, \Sigma)$. Partition $\boldsymbol{x}$ into two sub-vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, and partition the mean vector and covariance matrix accordingly as

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix} \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

(a) Define $\boldsymbol{z}_2 = \boldsymbol{x}_2 - \left[\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1) \right]$. Show that $\mathbb{E}\boldsymbol{z}_2 = 0$.

(b) Show that $\mathrm{Cov}(\boldsymbol{x}_1, \boldsymbol{z}_2) = \boldsymbol{0}$ and $\mathrm{Var}(\boldsymbol{z}_2) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$.

(c) Conclude from part (a) and (b) that

$$\mathbb{E}(\boldsymbol{x}_2 | \boldsymbol{x}_1) = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)$$

$$\mathrm{Var}(\boldsymbol{x}_2 | \boldsymbol{x}_1) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}.$$

(d) From now on we will focus on the covariance matrices, so we will assume that $\boldsymbol{\mu} = \boldsymbol{0}$ for simplicity. From part (b) we know that

$$\mathrm{Cov}\begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{z}_2 \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}. \end{pmatrix}$$

Now use the fact

$$\begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \Sigma_{21}\Sigma_{11}^{-1} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{z}_2 \end{pmatrix}$$

to verify that

$$\Sigma = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \Sigma_{21}\Sigma_{11}^{-1} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix} \begin{pmatrix} \boldsymbol{I} & \Sigma_{11}^{-1}\Sigma_{12} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix},$$

and

$$\Sigma^{-1} = \begin{pmatrix} \boldsymbol{I} & -\Sigma_{11}^{-1}\Sigma_{12} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\Sigma_{21}\Sigma_{11}^{-1} & \boldsymbol{I} \end{pmatrix}$$

(e) Let $\Theta = \Sigma^{-1}$ be the precision matrix, and denote its elements by θ_{jk}. Show that for any two components x_j and x_k of $\boldsymbol{x}$, it holds that

$$\mathrm{Cor}\left(x_j, x_k | \boldsymbol{x} \setminus \{x_j, x_k\}\right) = \frac{-\theta_{jk}}{\sqrt{\theta_{jj}\theta_{kk}}}.$$

This is also known as the *partial correlation* of x_j and x_k. This result tells us whether x_j and x_k are conditional independent given the rest of $\boldsymbol{x}$ depends on whether θ_{jk} is zero.

5.   (a) Use the graphical lasso to estimate the undirected graph for the `gene.txt` data. Try different choices of the tuning parameter λ. Report and plot the undirected graphs you found.

(b) (This part is a bonus question.) What are the corresponding BIC of the models you found?