2. (a) Ex 17.3 in "Elements of Statistical Learning".

(b) Use the results in (a) to derive the formula for the simple three variable case in lecture:

$$\rho_{AB\cdot C} \equiv \rho(A,B|C) = \frac{\rho_{AB} - \rho_{AC}\rho_{BC}}{\sqrt{1-\rho_{AC}^2}\sqrt{1-\rho_{BC}^2}}.$$

(c) Neighborhood selection considers regression of X_j on X_k, for all $k \neq j$ as discussed in lecture. Show the regression coefficient for neighborhood selection is related to the j,k element of Θ: $\beta_k^j = -\frac{\Theta_{jk}}{\Theta_{jj}}$.

Ex. 17.3 Let $\mathbf{\Sigma}$ be the covariance matrix of a set of p variables X. Consider the partial covariance matrix $\mathbf{\Sigma}_{a.b} = \mathbf{\Sigma}_{aa} - \mathbf{\Sigma}_{ab}\mathbf{\Sigma}_{bb}^{-1}\mathbf{\Sigma}_{ba}$ between the two subsets of variables $X_a = (X_1, X_2)$ consisting of the first two, and X_b the rest. This is the covariance matrix between these two variables, after linear adjustment for all the rest. In the Gaussian distribution, this is the covariance matrix of the conditional distribution of $X_a|X_b$. The partial correlation coefficient $\rho_{jk|\text{rest}}$ between the pair X_a conditional on the rest X_b, is simply computed from this partial covariance. Define $\mathbf{\Theta} = \mathbf{\Sigma}^{-1}$.

1. Show that $\mathbf{\Sigma}_{a.b} = \mathbf{\Theta}_{aa}^{-1}$.
2. Show that if any off-diagonal element of $\mathbf{\Theta}$ is zero, then the partial correlation coefficient between the corresponding variables is zero.
3. Show that if we treat $\mathbf{\Theta}$ as if it were a covariance matrix, and compute the corresponding "correlation" matrix

$$\mathbf{R} = \text{diag}(\mathbf{\Theta})^{-1/2} \cdot \mathbf{\Theta} \cdot \text{diag}(\mathbf{\Theta})^{-1/2}, \quad (17.40)$$

then $r_{jk} = -\rho_{jk|\text{rest}}$

(a) 1, $\Sigma \Theta = I \Rightarrow \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} \begin{pmatrix} \Theta_{aa} & \Theta_{ab} \\ \Theta_{ba} & \Theta_{bb} \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \Rightarrow \begin{cases} \Sigma_{aa}\Theta_{aa} + \Sigma_{ab}\Theta_{ba} = I & ① \\ \Sigma_{ba}\Theta_{aa} + \Sigma_{bb}\Theta_{ba} = 0 & ② \end{cases}$

from ②: $\Sigma_{bb}\Theta_{ba} = -\Sigma_{ba}\Theta_{aa} \Rightarrow \Theta_{ba} = -\Sigma_{bb}^{-1}\Sigma_{ba}\Theta_{aa}$, substitude into ①: $\Sigma_{aa}\Theta_{aa} + \Sigma_{ab}(-\Sigma_{bb}^{-1}\Sigma_{ba}\Theta_{aa}) = I$

$\Rightarrow (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})\Theta_{aa} = I \Rightarrow \Sigma_{a.b}\Theta_{aa} = I \Rightarrow \Sigma_{a.b} = \Theta_{aa}^{-1}$

2, Let $X_a = (X_i, X_j)$, rest are given. here $\Theta_{aa} = \begin{pmatrix} \theta_{ii} & \theta_{ij} \\ \theta_{ji} & \theta_{jj} \end{pmatrix}$

So $\Sigma_{a.b} = \Theta_{aa}^{-1} = \frac{1}{\det(\Theta_{aa})}\begin{pmatrix} \theta_{jj} & -\theta_{ij} \\ -\theta_{ji} & \theta_{ii} \end{pmatrix}$, and $\det(\Theta_{aa}) = \theta_{ii}\theta_{jj} - \theta_{ij}^2$

$Cov(X_i, X_j | rest) = \frac{-\theta_{ij}}{\det(\Theta_{aa})}$, if $\theta_{ij} = 0 \Rightarrow cov(X_i, X_j|rest) = 0 \Rightarrow \rho_{ij}|rest = 0$

3, $\Sigma_{a.b} = \frac{1}{\det(\Theta_{aa})}\begin{pmatrix} \theta_{jj} & -\theta_{ij} \\ -\theta_{ji} & \theta_{ii} \end{pmatrix} \Rightarrow \rho_{ij}|rest = \frac{-\theta_{ij}/\det(\Theta_{aa})}{\sqrt{(\theta_{jj}/\det(\Theta_{aa}))\cdot(\theta_{ii}/\det(\Theta_{aa}))}} = \frac{-\theta_{ij}}{\sqrt{\theta_{jj}\theta_{ii}}}$

for R: $r_{ij} = (\theta_{ii})^{-\frac{1}{2}}\theta_{ij}(\theta_{jj})^{-\frac{1}{2}} \Rightarrow r_{ij} = -\rho_{ij}|rest$

(b) Suppose ABC already normalized: $\sigma=1, \mu=0$ (for normalization doesn't affect correlation)

then correlation matrix $\Sigma = \begin{pmatrix} 1 & \rho_{AB} & \rho_{AC} \\ \rho_{AB} & 1 & \rho_{BC} \\ \rho_{AC} & \rho_{BC} & 1 \end{pmatrix}$, then for θ_{ij} in $\Theta = \Sigma^{-1}$, $\theta_{ij} = \frac{(-1)^{i+j} M_{ji}}{\det(\Sigma)}$

for θ_{12}: $M_{21} = \begin{vmatrix} \rho_{AB} & \rho_{AC} \\ \rho_{BC} & 1 \end{vmatrix} = \rho_{AB} - \rho_{AC}\rho_{BC} \Rightarrow \theta_{12} = \frac{\rho_{AC}\rho_{BC} - \rho_{AB}}{\det(\Sigma)}$

θ_{11}: $M_{11} = \begin{vmatrix} 1 & \rho_{BC} \\ \rho_{BC} & 1 \end{vmatrix} = 1 - \rho_{BC}^2 \Rightarrow \theta_{11} = \frac{1-\rho_{BC}^2}{\det(\Sigma)}$, and so on, $\theta_{22} = \frac{1-\rho_{AC}^2}{\det(\Sigma)}$

according to conclusion in a.3., $\rho_{AB\cdot C} = \rho(A,B|C) = \frac{-\theta_{12}}{\sqrt{\theta_{11}\theta_{22}}} = \frac{\rho_{AB} - \rho_{AC}\rho_{BC}}{\sqrt{1-\rho_{BC}^2}\cdot\sqrt{1-\rho_{AC}^2}}$

(c) $X \sim N(0, \Sigma)$, $p(x) \propto \exp(-\frac{1}{2}x^T\Theta x)$, $\Theta = \Sigma^{-1}$

$x^T\Theta x = \sum_a \sum_b x_a\theta_{ab}x_b = x_j\Theta_{jj}x_j + \sum_{k\neq j} x_j\Theta_{jk}x_k + \sum_{k\neq j} x_k\Theta_{kj}x_j +$ rest without $x_j \Rightarrow x^T\Theta x = \Theta_{jj}x_j^2 + 2x_j\sum_{k\neq j}\Theta_{jk}x_k +$ rest $(*)$

As for $p(X_j|rest)$, it's still Gaussian $\Rightarrow p(X_j|rest) \propto \exp[-\frac{1}{2\sigma_{j|rest}^2}(x_j - \mu_{j|rest})^2] = \exp[-\frac{1}{2\sigma_{j|rest}^2}x_j^2 - \frac{\mu_{j|rest}}{\sigma_{j|rest}^2}x_j + rest]$

compared with $\exp(-\frac{1}{2}(*))$: there are $\frac{1}{\sigma_{j|rest}^2} = \Theta_{jj}$ and $\frac{\mu_{j|rest}}{\sigma_{j|rest}^2} = -\sum_{k\neq j}\Theta_{jk}x_k \Rightarrow \Theta_{jj}\cdot\mu_{j|rest} = -\sum_{k\neq j}\Theta_{jk}x_k$

$\Rightarrow \mu_{j|rest} = E[X_j|rest] = \sum_{k\neq j}(-\frac{\Theta_{jk}}{\Theta_{jj}})x_k \Rightarrow \beta_k^j = -\frac{\Theta_{jk}}{\Theta_{jj}}$