# Tuesday Precept 7: LASSO and Graphic LASSO Regression

Oct 29, 2024

*Lecturer: Qishuo Yin* *Scribe: Qishuo Yin*

## 1 Precision Matrix

- **Definition**
  Suppose $\mathbf{X} = (X_1, \cdots, X_d)$ a $d$-dimenional Gaussian vector with mean vector $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_d)$ and covariance matrix $\Sigma$, i.e. $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$, we partition the vector into two parts $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ with $\mathbf{X}_1 = (X_1, \cdots, X_{d_1})$ and $\mathbf{X}_2 = (X_{d_1+1}, \cdots, X_d)$ with $d_1 + d_2 = d$, $\boldsymbol{\mu}_1 = (\mu_1, \cdots, \mu_{d_1})$, $\boldsymbol{\mu}_2 = (\mu_{d_1+1}, \cdots, \mu_d)$, and

$$\Sigma = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix}.$$

  The precision matrix $\boldsymbol{P}$ (also written as $\boldsymbol{\Omega}$ sometimes) is defined as

$$\boldsymbol{P} = \boldsymbol{\Sigma^{-1}} = \begin{bmatrix} P_{1,1} & P_{1,2} \\ P_{2,1} & P_{2,2} \end{bmatrix}.$$

- **Properties of covariance matrix**
  $\Sigma_{2,1} = \Sigma_{1,2}^t$
  $\mathbf{X}_1 \sim N_{d_1}(\boldsymbol{\mu}_1, \Sigma_{1,1})$
  $\mathbf{X}_2 \sim N_{d_2}(\boldsymbol{\mu}_2, \Sigma_{2,2})$

- **Properties of precision matrix**
  $P_{1,1}^{-1} = \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}$
  $P_{1,2} = P_{2,1}^t = -P_{1,1}\Sigma_{1,2}\Sigma_{2,2}^{-1}$
  $P_{2,2}^{-1} = \Sigma_{2,2} - \Sigma_{2,1}P_{1,1}^{-1}\Sigma_{1,2}$.

- **Proof** [Lemma] By Schur complement, the inverse of $2 \times 2$ block matrix:

$$S := \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

  If $A^{-1}$ or $D^{-1}$ exist, we know that matrix $S$ can be inverted.

$$S^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix}$$

Then, we use the lemma in our proof for the precision matrix:

$$
\begin{aligned}
\boldsymbol{P} &= \boldsymbol{\Sigma^{-1}} \\
&= \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix}^{-1} \\
&= \begin{bmatrix} (\Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1})^{-1} & -(\Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1})^{-1}\Sigma_{1,2}\Sigma_{2,2}^{-1} \\ -\Sigma_{2,2}^{-1}\Sigma_{2,1}(\Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1})^{-1} & (\Sigma_{2,2} - \Sigma_{2,1}\Sigma_{1,1}^{-1}\Sigma_{1,2})^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \Sigma_{1,1}^{-1} - \Sigma_{2,1}^{-1}\Sigma_{2,2}\Sigma_{1,2}^{-1} & -\Sigma_{1,1}\Sigma_{1,2}\Sigma_{2,2}^{-1} + \Sigma_{2,1}^{-1}\Sigma_{2,2}\Sigma_{1,2}^{-1}\Sigma_{1,2}\Sigma_{2,2}^{-1} \\ -\Sigma_{2,2}^{-1}\Sigma_{2,1}\Sigma_{1,1}^{-1} + \Sigma_{2,2}^{-1}\Sigma_{2,1}\Sigma_{2,1}^{-1}\Sigma_{2,2}\Sigma_{1,2}^{-1} & \Sigma_{2,2}^{-1} - \Sigma_{1,2}^{-1}\Sigma_{1,1}\Sigma_{2,1}^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \Sigma_{1,1}^{-1} - \Sigma_{2,1}^{-1}\Sigma_{2,2}\Sigma_{1,2}^{-1} & -\Sigma_{1,1}\Sigma_{1,2}\Sigma_{2,2}^{-1} + \Sigma_{2,1}^{-1} \\ -\Sigma_{2,2}^{-1}\Sigma_{2,1}\Sigma_{1,1}^{-1} + \Sigma_{1,2}^{-1} & \Sigma_{2,2}^{-1} - \Sigma_{1,2}^{-1}\Sigma_{1,1}\Sigma_{2,1}^{-1} \end{bmatrix}
\end{aligned}
$$

Then

$$
\begin{aligned}
P_{1,1}^{-1} &= \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1} \\
P_{1,2} &= -(\Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1})^{-1}\Sigma_{1,2}\Sigma_{2,2}^{-1} \\
&= P_{2,1}^{t} = -P_{1,1}\Sigma_{1,2}\Sigma_{2,2}^{-1} \\
P_{2,2}^{-1} &= \Sigma_{2,2} - \Sigma_{2,1}P_{1,1}^{-1}\Sigma_{1,2}
\end{aligned}
$$

- **Conditional distribution**
  The conditional distribution of the random vector $\mathbf{X}_1$ given $\mathbf{X}_2$, is still a Gaussian distribution, and its mean vector can be given as

$$\boldsymbol{\mu}_{\mathbf{X}_1|\mathbf{X_2}} = \boldsymbol{\mu}_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(X_2 - \boldsymbol{\mu}_2)$$

  and its covariance matrix can be given as

$$\Sigma_{\mathbf{X}_1|\mathbf{X_2}} = \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}$$

- **Intuitive understanding of precision matrix**
  $P_{i,j} = 0$ is equivalent to $X_i$ and $X_j$ are conditionally indep, given rest variables. An example can be given by section 9.4 of [1]

# 2 Graphic LASSO Regression

The Graphical LASSO aims to propose an algorithm to find a sparse estimate of the inverse of a large covariance matrix. It adds a penalized term to the likelihood function to achieve the optimization:

$$\hat{S} = \arg\inf_{S} \left(\text{trace}(\hat{P}S) - \log|P| + \rho\|P\|_1\right)$$

$$
\boldsymbol{\Omega} = \begin{bmatrix} 1 & 0.2 & 0.2 & 0.2 & 0 & 0 \\ 0.2 & 1 & 0 & 0 & 0 & 0 \\ 0.2 & 0 & 1 & 0 & 0.2 & 0.2 \\ 0.2 & 0 & 0 & 1 & 0.2 & 0.2 \\ 0 & 0 & 0.2 & 0.2 & 1 & 0 \\ 0 & 0 & 0.2 & 0.2 & 0 & 1 \end{bmatrix}
$$

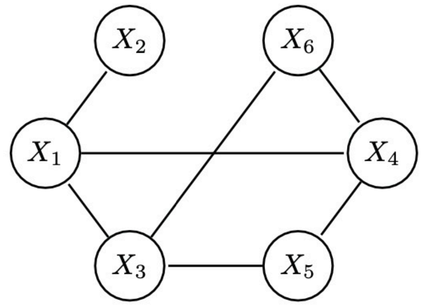


Figure 9.4: The left panel is a precision matrix and the right panel is its graphical representation. Note that only the zero pattern of the precision matrix is used in the graphical representation.

Figure 1: Caption

where

$$
S = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}})(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}})^T
$$

Meanwhile, as an extension, Zhang and Zou (2014) introduced an empirical loss minimization framework to the precision matrix estimation problem, which could be a generalized version of the graphic LASSO. This method solves the following optimization:

$$
\hat{P} = \arg\min_{P>0} \left( L(\Sigma, P) + \lambda \sum_{i \neq j} |P_{i,j}| \right)
$$

where L is a proper loss function that satisfies two conditions:
(i) $L)(\Sigma, P)$ is a smooth convex function of $\Sigma$
(ii) the unique minimizer of $L(\Sigma, P)$ occurs at the true $\Sigma^*$, $P$ at true $\Sigma^*$. If you have interest, you can also see Constrained L1-minimization for Inverse Matrix Estimation and Danzig selector for reference.

# References

[1] J. Fan, R. Li, C.-H. Zhang, and H. Zou. *Statistical foundations of data science*. Chapman and Hall/CRC, 2020.