Selected Topics in Mathematics of Learning

High-Dimensional Statistics

Lecturer: Marius Yamakou

Winter Semester 2024/25
Department of Data Science, FAU

December 3, 2024

Part V: Large Inverse Covariance Matrices

Objectives:

1. **Understand the Concept of the Precision Matrix**
   - Define the precision matrix and explain its relationship to the inverse of the covariance matrix.
   - Explore its role in capturing conditional independencies in multivariate distributions.

Part V: Large Inverse Covariance Matrices

Objectives:

1. **Understand the Concept of the Precision Matrix**
   - Define the precision matrix and explain its relationship to the inverse of the covariance matrix.
   - Explore its role in capturing conditional independencies in multivariate distributions.
2. **Analyze Graphical Representations of Relationships in Data**
   - Interpret the structure of graphical models and their connection to the precision matrix.
   - Distinguish between independence and conditional independence in the context of graphs.

Part V: Large Inverse Covariance Matrices

Objectives:

1. **Understand the Concept of the Precision Matrix**
   - Define the precision matrix and explain its relationship to the inverse of the covariance matrix.
   - Explore its role in capturing conditional independencies in multivariate distributions.
2. **Analyze Graphical Representations of Relationships in Data**
   - Interpret the structure of graphical models and their connection to the precision matrix.
   - Distinguish between independence and conditional independence in the context of graphs.
3. **Deepen Knowledge of the Multivariate Normal Distribution**
   - Examine the role of the precision matrix within the multivariate normal framework.
   - Relate matrix theory concepts to covariance and precision matrices.

4. **Evaluate Methods for Estimating the Precision Matrix**
   - Discuss the limitations of directly using $\widehat{\Sigma}^{-1}$ (the inverse of the sample covariance matrix).
   - Introduce and explain the graphical LASSO as a sparsity-inducing method for precision matrix estimation.

4. **Evaluate Methods for Estimating the Precision Matrix**
   - Discuss the limitations of directly using $\widehat{\Sigma}^{-1}$ (the inverse of the sample covariance matrix).
   - Introduce and explain the graphical LASSO as a sparsity-inducing method for precision matrix estimation.
5. **Interpret and Analyze Data Using Precision Matrices**
   - Investigate real-world examples of sparsity in precision matrices.
   - Compare and contrast covariance matrices with precision matrices in practical contexts.

4. **Evaluate Methods for Estimating the Precision Matrix**
   - Discuss the limitations of directly using $\widehat{\Sigma}^{-1}$ (the inverse of the sample covariance matrix).
   - Introduce and explain the graphical LASSO as a sparsity-inducing method for precision matrix estimation.
5. **Interpret and Analyze Data Using Precision Matrices**
   - Investigate real-world examples of sparsity in precision matrices.
   - Compare and contrast covariance matrices with precision matrices in practical contexts.
6. **Explore Regularization and Model Selection Techniques**
   - Understand the role of regularization in estimating sparse precision matrices.
   - Learn criteria and strategies for selecting the optimal regularization parameter.

Outline

1. Precision matrix
   1. Graphs
   2. Independence vs uncorrelatedness
   3. Independence vs conditional independence
   4. Multivariate normal
   5. Some more matrix theory

Outline

Outline

Outline

1. Precision Matrix

What is the Precision Matrix?

- The precision matrix, denoted by: $\Theta = (\Theta_{ij})_{i,j=1,\ldots,p} = \Sigma^{-1}$ is the inverse of the covariance matrix.
- Many statistical procedures focus on estimating Θ rather than Σ.
- Why? Θ reveals **conditional independence relationships**, providing insights into the underlying structure of the data.

Why is the Precision Matrix Important?

- It helps uncover relationships between variables by describing how they depend on one another after accounting for all other variables.
- This makes Θ a fundamental tool in learning data structures and interpreting dependencies.

To interpret the precision matrix properly, we will need to learn more about

- **Undirected Graphs**
- **Independence vs Conditional Independence**
- **Multivariate Normal Distribution**
- **Matrix Theory**

1.1 Undirected Graph

- A **graph** G consists of:
  - **Vertices** (V): A set of points (nodes).
  - **Edges** (E): A set of pairs of vertices, representing connections between nodes.

1.1 Undirected Graph

- A **graph** G consists of:
  - **Vertices** (V): A set of points (nodes).
  - **Edges** (E): A set of pairs of vertices, representing connections between nodes.
- **Adjacency**:
  - Two vertices X and Y are adjacent if there is an edge between them.

1.1 Undirected Graph

- A **graph** G consists of:
  - **Vertices** (V): A set of points (nodes).
  - **Edges** (E): A set of pairs of vertices, representing connections between nodes.
- **Adjacency**:
  - Two vertices X and Y are adjacent if there is an edge between them.
- **Path**:
  - A sequence of vertices $X_1, X_2, \ldots, X_n$, where each pair (X_i, X_{i+1}) is connected by an edge.

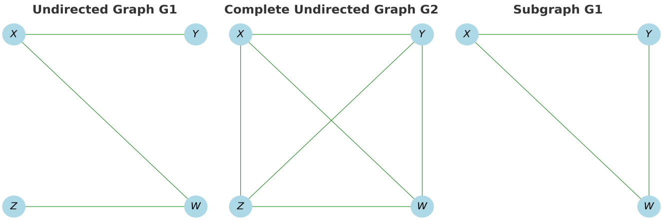
1.1 Undirected Graph

- A **graph** G consists of:
  - **Vertices** (V): A set of points (nodes).
  - **Edges** (E): A set of pairs of vertices, representing connections between nodes.
- **Adjacency**:
  - Two vertices X and Y are adjacent if there is an edge between them.
- **Path**:
  - A sequence of vertices $X_1, X_2, \ldots, X_n$, where each pair (X_i, X_{i+1}) is connected by an edge.
- **Complete Graph**:
  - A graph where every pair of vertices is connected by an edge.

1.1 Undirected Graph

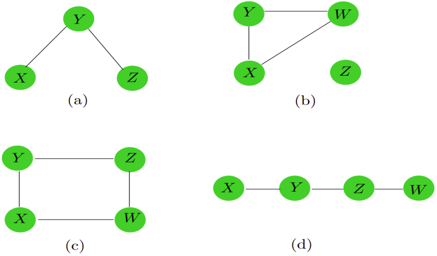
- A **graph** G consists of:
  - **Vertices** (V): A set of points (nodes).
  - **Edges** (E): A set of pairs of vertices, representing connections between nodes.
- **Adjacency**:
  - Two vertices X and Y are adjacent if there is an edge between them.
- **Path**:
  - A sequence of vertices $X_1, X_2, \ldots, X_n$, where each pair (X_i, X_{i+1}) is connected by an edge.
- **Complete Graph**:
  - A graph where every pair of vertices is connected by an edge.
- **Subgraph**:
  - A subset $U \subseteq V$, along with all edges between vertices in U.

1.1 Undirected Graph



Notice that G1 is a subgraph to graph G2 but not to graph G1.

1.1 Undirected Graph



The adjacency matrix A corresponding to the graph in (b) is given by:

$$A = \begin{bmatrix} xx & xy & xz & xw \\ yx & yy & yz & yw \\ zx & zy & zz & zw \\ wx & wy & wz & ww \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

1.2 Independence vs Uncorrelatedness

Two variables X and Y are **independent** if their joint distribution factorizes into the product of their so-called **marginal distributions**.

1.2 Independence vs Uncorrelatedness

Two variables X and Y are **independent** if their joint distribution factorizes into the product of their so-called **marginal distributions**.

For **Continuous** independent random variables (denoted $X \perp\!\!\!\perp Y$):

we have: $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$

where $\quad F_{X,Y}(x,y) = \mathbb{P}(X \leq x, Y \leq y) \quad$ and $\quad f_{X,Y}(x,y) = \dfrac{\partial^2 F_{X,Y}(x,y)}{\partial x \, \partial y}$

1.2 Independence vs Uncorrelatedness

Two variables X and Y are **independent** if their joint distribution factorizes into the product of their so-called **marginal distributions**.

For **Continuous** independent random variables (denoted $X \perp\!\!\!\perp Y$):

we have: $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$

where $\quad F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y) \quad$ and $\quad f_{X,Y}(x, y) = \dfrac{\partial^2 F_{X,Y}(x, y)}{\partial x \, \partial y}$

The marginal distribution of a single variable is obtained by summing or integrating out the other variable(s) from the joint distribution:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dy \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dx$$

In terms of events we can formulate: Two events A and B are **independent** if and only if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

1.2 Independence vs Uncorrelatedness

Two variables X and Y are **independent** if their joint distribution factorizes into the product of their so-called **marginal distributions**.

For **Continuous** independent random variables (denoted $X \perp\!\!\!\perp Y$):

we have: $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$

$$\text{where} \quad F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y) \quad \text{and} \quad f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \, \partial y}$$

The marginal distribution of a single variable is obtained by summing or integrating out the other variable(s) from the joint distribution:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx$$

In terms of events we can formulate: Two events A and B are **independent** if and only if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

Example:

- Rolling a die: Getting a 6 on the first roll (A) and getting a 6 on the second roll (B) are independent events.
- By contrast, the event of getting a 6 the first time a die is rolled and the event that the sum of the numbers seen on the first and second trial is 8 are not independent.

1.2 Independence vs Uncorrelatedness

Uncorrelatedness does not imply independence.

1.2 Independence vs Uncorrelatedness

Uncorrelatedness does not imply independence.

Example:

- Let W be a random variable such that:

$$\mathbb{P}(W = 1) = \mathbb{P}(W = -1) = \frac{1}{2}.$$

- Let $X \sim N(0, 1)$ (standard normal random variable).
- Define $Y := WX$.

1.2 Independence vs Uncorrelatedness

Uncorrelatedness does not imply independence.

Example:

- Let W be a random variable such that:

$$\mathbb{P}(W = 1) = \mathbb{P}(W = -1) = \frac{1}{2}.$$

- Let $X \sim N(0, 1)$ (standard normal random variable).
- Define $Y := WX$.

Analysis:

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathbb{E}[XY] = \mathbb{E}[X \cdot WX] = \mathbb{E}[X^2 W] \\
&= \mathbb{E}[X^2] \cdot \mathbb{E}[W] && \text{(since } X^2 \text{ and } W \text{ are independent)} \\
&= 1 \cdot \mathbb{E}[W].
\end{aligned}$$

1.2 Independence vs Uncorrelatedness

Uncorrelatedness does not imply independence.

Example:

- Let W be a random variable such that:

$$\mathbb{P}(W=1)=\mathbb{P}(W=-1)=\frac{1}{2}.$$

- Let $X \sim N(0,1)$ (standard normal random variable).
- Define $Y := WX$.

Analysis:

$$\begin{aligned}\text{Cov}(X,Y) &= \mathbb{E}[XY]=\mathbb{E}[X\cdot WX]=\mathbb{E}[X^2W]\\ &= \mathbb{E}[X^2]\cdot\mathbb{E}[W] && \text{(since } X^2 \text{ and } W \text{ are independent)}\\ &= 1\cdot\mathbb{E}[W].\end{aligned}$$

Calculation of $\mathbb{E}[W]$: $\mathbb{E}[W]=1\cdot\mathbb{P}(W=1)+(-1)\cdot\mathbb{P}(W=-1)=\frac{1}{2}-\frac{1}{2}=0.$

1.2 Independence vs Uncorrelatedness

Uncorrelatedness does not imply independence.

Example:

- Let W be a random variable such that:

$$\mathbb{P}(W = 1) = \mathbb{P}(W = -1) = \frac{1}{2}.$$

- Let $X \sim N(0, 1)$ (standard normal random variable).
- Define $Y := WX$.

Analysis:

$$\begin{aligned} \mathrm{Cov}(X, Y) &= \mathbb{E}[XY] = \mathbb{E}[X \cdot WX] = \mathbb{E}[X^2 W] \\ &= \mathbb{E}[X^2] \cdot \mathbb{E}[W] && \text{(since } X^2 \text{ and } W \text{ are independent)} \\ &= 1 \cdot \mathbb{E}[W]. \end{aligned}$$

Calculation of $\mathbb{E}[W]$: $\mathbb{E}[W] = 1 \cdot \mathbb{P}(W = 1) + (-1) \cdot \mathbb{P}(W = -1) = \frac{1}{2} - \frac{1}{2} = 0.$
Conclusion: $\mathrm{Cov}(X, Y) = 0$. Thus, although $\mathrm{Cov}(X, Y) = 0$, X and Y are **not** independent because Y is defined as a function of X ($Y = WX$).

1.3 Independence vs conditional independence

Two variables X and Y are conditionally independent given variable Z, if and only if their conditional distribution factorizes as:

$$f_{X,Y|Z=z}(x,y) = f_{X|Z=z}(x)f_{Y|Z=z}(y)$$

$$f_{X|Z=z}(x) = \frac{f_{X,Z}(x,z)}{f_Z(z)}$$

1.3 Independence vs conditional independence

Two variables X and Y are conditionally independent given variable Z, if and only if their conditional distribution factorizes as:

$$f_{X,Y|Z=z}(x,y) = f_{X|Z=z}(x) f_{Y|Z=z}(y)$$

$$f_{X|Z=z}(x) = \frac{f_{X,Z}(x,z)}{f_Z(z)}$$

Notation: $X \perp\!\!\!\perp Y | Z$. (For continuous random variables.)
Let A, B and C be events. Then, A and B are conditionally independent given C, if and only if $P(C) > 0$ and ,

$$\mathbb{P}(A \cap B | C) = \mathbb{P}(A|C)\mathbb{P}(B|C).$$

1.3 Independence vs conditional independence

Two variables X and Y are conditionally independent given variable Z, if and only if their conditional distribution factorizes as:

$$f_{X,Y|Z=z}(x,y) = f_{X|Z=z}(x) f_{Y|Z=z}(y)$$

$$f_{X|Z=z}(x) = \frac{f_{X,Z}(x,z)}{f_Z(z)}$$

Notation: $X \perp\!\!\!\perp Y | Z$. (For continuous random variables.)
Let A, B and C be events. Then, A and B are conditionally independent given C, if and only if $P(C) > 0$ and ,

$$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C)\mathbb{P}(B|C).$$

In that case we have $\mathbb{P}(A|B, C) = \mathbb{P}(A|C)$, i.e. in light of information C, B provides no (further) information about A.

$$\mathbb{P}(A \mid C) = \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)}$$

1.3 Independence vs conditional independence

Example:

- **Key Point:** Conditional independence depends on the nature of the third event.

1.3 Independence vs conditional independence

Example:

- **Key Point:** Conditional independence depends on the nature of the third event.
- **Independent Events:**
  - Roll two dice: The results of the two dice are independent.
  - Observing the result of one die gives no information about the result of the other die.

1.3 Independence vs conditional independence

Example:

- **Key Point:** Conditional independence depends on the nature of the third event.
- **Independent Events:**
  - Roll two dice: The results of the two dice are independent.
  - Observing the result of one die gives no information about the result of the other die.
- **Conditionally Dependent Events:**
  - Suppose the result of the first die is 3.

1.3 Independence vs conditional independence

Example:

- **Key Point:** Conditional independence depends on the nature of the third event.
- **Independent Events:**
  - Roll two dice: The results of the two dice are independent.
  - Observing the result of one die gives no information about the result of the other die.
- **Conditionally Dependent Events:**
  - Suppose the result of the first die is 3.
  - Someone tells you a third event: "The sum of the two results is even."

1.3 Independence vs conditional independence

Example:

- **Key Point:** Conditional independence depends on the nature of the third event.
- **Independent Events:**
  - Roll two dice: The results of the two dice are independent.
  - Observing the result of one die gives no information about the result of the other die.
- **Conditionally Dependent Events:**
  - Suppose the result of the first die is 3.
  - Someone tells you a third event: "The sum of the two results is even."
  - This extra information restricts the possible outcomes of the second die to odd numbers.

1.3 Independence vs conditional independence

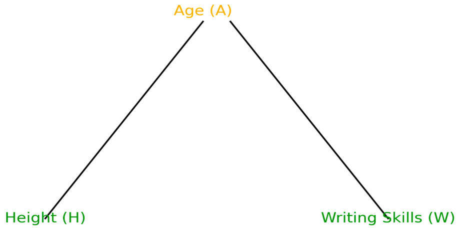
Example:

- **Key Point:** Conditional independence depends on the nature of the third event.
- **Independent Events:**
  - Roll two dice: The results of the two dice are independent.
  - Observing the result of one die gives no information about the result of the other die.
- **Conditionally Dependent Events:**
  - Suppose the result of the first die is 3.
  - Someone tells you a third event: "The sum of the two results is even."
  - This extra information restricts the possible outcomes of the second die to odd numbers.
- **Conclusion:** While the two dice are independent, they are **not conditionally independent** given the sum is even.

1.3 Independence vs conditional independence: Conditioning Example

Observations:

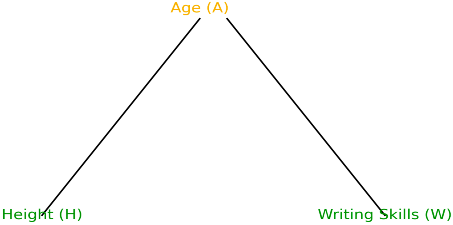
- Random sample of school kids.
- Age (A) correlates with both height (H) and writing skills (W).



1.3 Independence vs conditional independence: Conditioning Example

Observations:

- Random sample of school kids.
- Age (A) correlates with both height (H) and writing skills (W).



Conditioning:

- Fix $A = 10$ (e.g., consider only one age group).
- **Effect:** Removes correlation between H and W that is due to age as a common factor.

1.3 Independence vs conditional independence: Conditioning Example

Conditional Probability:

$$P(H \cap W \mid A = 10) = P(H \mid A = 10)P(W \mid A = 10)$$

1.3 Independence vs conditional independence: Conditioning Example

Conditional Probability:

$$P(H \cap W \mid A = 10) = P(H \mid A = 10)P(W \mid A = 10)$$

- **Assumption:** H and W are conditionally independent given A.

1.3 Independence vs conditional independence: Conditioning Example

Conditional Probability:

$$P(H \cap W \mid A = 10) = P(H \mid A = 10)P(W \mid A = 10)$$

- **Assumption:** H and W are conditionally independent given A.

Example:

$$P(H \leq 170 \cap W \leq 80 \mid A = 10) = P(H \leq 170 \mid A = 10)P(W \leq 80 \mid A = 10)$$

1.3 Independence vs conditional independence: Conditioning Example

Conditional Probability:

$$P(H \cap W \mid A = 10) = P(H \mid A = 10)P(W \mid A = 10)$$

- **Assumption:** H and W are conditionally independent given A.

Example:

$$P(H \leq 170 \cap W \leq 80 \mid A = 10) = P(H \leq 170 \mid A = 10)P(W \leq 80 \mid A = 10)$$

Adding Another Variable:

$$P(H \cap W \mid V, A) = P(H \mid V, A)P(W \mid V, A)$$

1.3 Independence vs conditional independence: Conditioning Example

Conditional Probability:

$$P(H \cap W \mid A = 10) = P(H \mid A = 10)P(W \mid A = 10)$$

- **Assumption:** H and W are conditionally independent given A.

Example:

$$P(H \leq 170 \cap W \leq 80 \mid A = 10) = P(H \leq 170 \mid A = 10)P(W \leq 80 \mid A = 10)$$

Adding Another Variable:

$$P(H \cap W \mid V, A) = P(H \mid V, A)P(W \mid V, A)$$

- **Assumption:** H and W are conditionally independent given both V and A.

1.4 Multivariate normal

Uncorrelated and independent
Suppose that $X = (X_1, \dots, X_p)^\top \sim \mathcal{N}(\mu, \Sigma)$.
Partition the multivariate normal distribution:

$$X = \begin{pmatrix} Y_a \\ Y_b \end{pmatrix}$$

1.4 Multivariate normal

Uncorrelated and independent

Suppose that $X = (X_1, \dots, X_p)^\top \sim \mathcal{N}(\mu, \Sigma)$.
Partition the multivariate normal distribution:

$$X = \begin{pmatrix} Y_a \\ Y_b \end{pmatrix}$$

The two random vectors Y_a and Y_b are independent if and only if they are uncorrelated.

$$\mathrm{Cov}(X) = \mathbb{E}\left[\begin{pmatrix} Y_a \\ Y_b \end{pmatrix} \begin{pmatrix} Y_a' & Y_b' \end{pmatrix} \right] = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$$

Marginal distribution

Partition the multivariate normal distribution:

$$X = \begin{pmatrix} Y_a \\ Y_b \end{pmatrix} \quad \mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}.$$

1.4 Multivariate normal

Marginal distribution
Partition the multivariate normal distribution:

$$X = \begin{pmatrix} Y_a \\ Y_b \end{pmatrix} \quad \mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}.$$

Conveniently, the marginal distributions are exactly what you would intuitively think they should be:

$$Y_a \sim \mathcal{N}(\mu_a, \Sigma_{aa}).$$

1.4 Multivariate normal

Conditional

Partition the multivariate normal distribution:

$$X = \begin{pmatrix} Y_a \\ Y_b \end{pmatrix} \quad \mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}.$$

1.4 Multivariate normal

Conditional

Partition the multivariate normal distribution:

$$X = \begin{pmatrix} Y_a \\ Y_b \end{pmatrix} \quad \mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}.$$

A more complicated question: what is the distribution of Y_a given Y_b?

This gets messy if Σ is singular, but if Σ is full rank, then

$$Y_a | Y_b = y_b \sim \mathcal{N}(\mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(y_b - \mu_b), \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}),$$

where $\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} = \mathrm{Cov}(Y_a | Y_b = y_b)$.

1.4 Multivariate normal

Conditional

Partition the multivariate normal distribution:

$$X = \begin{pmatrix} Y_a \\ Y_b \end{pmatrix} \quad \mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}.$$

A more complicated question: what is the distribution of Y_a given Y_b?

This gets messy if Σ is singular, but if Σ is full rank, then

$$Y_a | Y_b = y_b \sim \mathcal{N}(\mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(y_b - \mu_b), \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}),$$

where $\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} = \mathrm{Cov}(Y_a | Y_b = y_b)$.

Note that if $\Sigma_{ab} = 0$, then Y_a and Y_b are independent and $Y_a | Y_b \sim \mathcal{N}(\mu_a, \Sigma_{aa})$.

1.5. Some more matrix theory

Inverse of a 2×2 block matrix

Suppose we have a block matrix

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$$

with $\mathbf{D}$ and $\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ invertible.

1.5. Some more matrix theory

Inverse of a 2×2 block matrix

Suppose we have a block matrix

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$$

with $\mathbf{D}$ and $\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ invertible. Then, the inverse of the matrix $\mathbf{M}$ is given by

$$\mathbf{M}^{-1} = \begin{bmatrix} \left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & -\left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}\left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix}.$$