Advanced Statistical Inference
Refresher on linear algebra

Maurizio Filippone
Maurizio.Filippone@eurecom.fr

Department of Data Science
EURECOM

Overview

Basic definitions and properties

- Vectors $\mathbf{v} = [v_i]$

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

- Matrices $\mathbf{A} = [a_{ij}]$

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Basic definitions and properties

- Matrix addition $\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}]$
- Scalar multiplication $\gamma \mathbf{B} = [\gamma\, b_{ij}]$
- Matrix-vector multiplication $\mathbf{Av} = \left[\sum_k a_{ik} v_k\right]$
- Matrix-matrix multiplication $\mathbf{AB} = \left[\sum_k a_{ik} b_{kj}\right]$

Basic definitions and properties

- Inverse $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$
- Transpose $\mathbf{A}^\top = [a_{ji}]$
- Transpose of products $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$
- Symmetric matrices $\mathbf{A} = \mathbf{A}^\top$
- Determinant $|\mathbf{A}|$ (on the board)
- Determinant $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$
- Determinant $|\mathbf{A}^\top| = |\mathbf{A}|$
- Trace $\mathrm{Tr}(\mathbf{A}) = \sum_k a_{kk}$
- Trace is permutation invariant $\mathrm{Tr}(\mathbf{ABC}) = \mathrm{Tr}(\mathbf{CAB}) = \mathrm{Tr}(\mathbf{BCA})$

Eigenvalues and eigenvectors

- Let $\mathbf{A} = [a_{i,j}]$ be an $n \times n$ matrix
- $\mathbf{A}$ is said to have an eigenvalue λ and (non-zero) eigenvector $\mathbf{x}$ corresponding to λ if

$$\mathbf{Ax} = \lambda \mathbf{x}.$$

- Eigenvalues are the λ values that solve the determinantal equation $|\mathbf{A} - \lambda \mathbf{I}| = 0$.

Spectral decomposition of symmetric matrices

- The Spectral decomposition theorem says that every square and symmetric matrix $\mathbf{A} = [a_{ij}]$ may be written

$$\mathbf{A} = \mathbf{CDC}^\top,$$

- The columns of $\mathbf{C}$ are the eigenvectors of $\mathbf{A}$
- The diagonal matrix $\mathbf{D}$ contains the corresponding eigenvalues

$$\mathbf{D} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

- The eigenvectors may be chosen to be orthonormal, so that $\mathbf{CC}^\top = \mathbf{C}^\top\mathbf{C} = \mathbf{I}$.
- Note the useful property: $|\mathbf{A}| = \prod_{i=1}^n \lambda_i$

Positive definite matrices

The $n \times n$ matrix $\mathbf{A}$ is said to be positive definite if

$$\mathbf{y}^\top \mathbf{A} \mathbf{y} > 0$$

for all $n \times 1$ vectors $\mathbf{y} \neq \mathbf{0}$.

Positive semidefinite matrices

The $n \times n$ matrix $\mathbf{A}$ is said to be positive semidefinite if

$$\mathbf{y}^\top \mathbf{A}\mathbf{y} \geq 0$$

for all $n \times 1$ vectors $\mathbf{y} \neq \mathbf{0}$.

Example: Show $\mathbf{X}^\top\mathbf{X}$ is positive semidefinite

Let $\mathbf{X}$ be an $n \times p$ matrix of real constants and $\mathbf{y}$ be $p \times 1$. Then $\mathbf{Z} = \mathbf{X}\mathbf{y}$ is $n \times 1$, and

$$\begin{aligned} & \mathbf{y}^\top (\mathbf{X}^\top\mathbf{X})\,\mathbf{y} \\ = \; & (\mathbf{X}\mathbf{y})^\top(\mathbf{X}\mathbf{y}) \\ = \; & \mathbf{Z}^\top\mathbf{Z} \\ = \; & \sum_{i=1}^{n} Z_i^2 \geq 0 \end{aligned}$$

Some properties of symmetric positive definite matrices

For a symmetric matrix,

Positive definite

$\Downarrow$

All eigenvalues positive

Showing Positive definite $\Rightarrow$ Eigenvalues positive

Let $\mathbf{A}$ be symmetric and positive definite.

- Spectral decomposition says $\mathbf{A} = \mathbf{C}\mathbf{D}\mathbf{C}^\top$.
- Using $\mathbf{y}^\top\mathbf{A}\mathbf{y} > 0$, let $\mathbf{y}$ be an eigenvector, say the third one.
- Because eigenvectors are orthonormal,

$$\begin{aligned} \mathbf{y}^\top\mathbf{A}\mathbf{y} &= \mathbf{y}^\top\mathbf{C}\mathbf{D}\mathbf{C}^\top\mathbf{y} \\ &= \begin{pmatrix} 0 & 0 & 1 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} \\ &= \lambda_3 \\ &> 0 \end{aligned}$$

Square root matrices

For symmetric, non-negative definite matrices

Define

$$\mathbf{D}^{1/2} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_n} \end{pmatrix}$$

So that

$$\begin{aligned} \mathbf{D}^{1/2}\mathbf{D}^{1/2} &= \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_n} \end{pmatrix}\begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_n} \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} = \mathbf{D} \end{aligned}$$

For a non-negative definite, symmetric matrix A

Define

$$\mathbf{A}^{1/2} = \mathbf{C}\mathbf{D}^{1/2}\mathbf{C}^\top$$

So that

$$\begin{aligned} \mathbf{A}^{1/2}\mathbf{A}^{1/2} &= \mathbf{C}\mathbf{D}^{1/2}\mathbf{C}^\top\mathbf{C}\mathbf{D}^{1/2}\mathbf{C}^\top \\ &= \mathbf{C}\mathbf{D}^{1/2}\,\mathbf{I}\,\mathbf{D}^{1/2}\mathbf{C}^\top \\ &= \mathbf{C}\mathbf{D}^{1/2}\mathbf{D}^{1/2}\mathbf{C}^\top \\ &= \mathbf{C}\mathbf{D}\mathbf{C}^\top \\ &= \mathbf{A} \end{aligned}$$

Cholesky decomposition

- Define lower triangular matrix $\mathbf{L}$

$$\mathbf{L} = \begin{bmatrix} L_{11} & 0 & \cdots & 0 \\ L_{21} & L_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ L_{n1} & L_{n2} & \cdots & L_{nn} \end{bmatrix}$$

  so that $\mathbf{A} = \mathbf{L}\mathbf{L}^\top$
- Cholesky algorithm computes $\mathbf{L}$ from $\mathbf{A}$
- $|\mathbf{A}| = |\mathbf{L}\mathbf{L}^\top| = |\mathbf{L}|^2 = \left(\prod_{i=1}^{n} L_{ii}\right)^2$