Statistical Learning

Tommaso Proietti

DEF Tor Vergata

Tools for data analysis

Vectors

A vector is a collection of numbers (scalars) ordered by column (or row).

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{bmatrix}, \quad \boldsymbol{a} = \begin{bmatrix} 5 \\ -2 \\ -3 \end{bmatrix}, \quad \boldsymbol{a} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

From a geometric perspective, $\boldsymbol{a}$ is a point in a n-dimensional space (a subset of $\mathbb{R}^n$), with coordinates provided by the elements a_i.

The symbol $\boldsymbol{a}'$ ($\boldsymbol{a}$ transpose) denotes a row vector: $\boldsymbol{a}' = [a_1, a_2, \ldots, a_n]$.
Special cases:
$\mathbf{0}$ (zero vector);
$\boldsymbol{e}_i = [0, \ldots, 0, 1, 0, \ldots, 0]'$ (unit vector);
$\boldsymbol{i} = [1, 1, \ldots, 1]'$.
Basic operations:

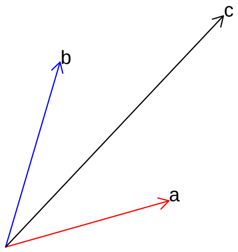
- Multiplication by a scalar.
  Let ρ denote a scalar; $\rho\boldsymbol{a}$ is the vector with elements $\{\rho a_i\}$. [Geom. Interpretation: a point in the same direction as $\boldsymbol{a}$].

- Sum of two vectors.
  Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be two vectors with the same size n; their sum $\boldsymbol{c} = \boldsymbol{a} + \boldsymbol{b}$ is the vector with elements $c_i = a_i + b_i$. [Geometric Int.: a new point in $\mathcal{R}^n$ obtained by the parallelogram rule].
- Linear combination.
  Let $\boldsymbol{a}$ and $\boldsymbol{b}$ denote 2 vectors and let ρ_1 e ρ_2 be two coefficients:

$$\rho_1 \boldsymbol{a} + \rho_2 \boldsymbol{b}$$

is their linear combination.

Figure: Sum of two vectors (parallelogram rule)



Inner (scalar) product

The inner product between the two n-dimensional vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is defined as

$$\boldsymbol{a}'\boldsymbol{b} = \sum_{i=1}^{n} a_i b_i$$

Example:

$$\boldsymbol{a} = \left[\begin{array}{r} 2 \\ -1 \\ 3 \end{array}\right], \boldsymbol{b} = \left[\begin{array}{r} 5 \\ -2 \\ -3 \end{array}\right], \Rightarrow \boldsymbol{a}'\boldsymbol{b} = 2 \cdot 5 + (-1) \cdot (-2) + 3 \cdot (-3) = 3$$

Example: $\bar{x} = n^{-1}\boldsymbol{i}'\boldsymbol{x}$ is the average of the elements of $\boldsymbol{x}$.

$$n = 4, \boldsymbol{x} = \begin{bmatrix} 3 \\ 4 \\ 2 \\ 7 \end{bmatrix}, \boldsymbol{i} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix},$$

$$\Rightarrow \frac{1}{4}\boldsymbol{x}'\boldsymbol{i} = \frac{1}{4}(3 \cdot 1 + 4 \cdot 1 + 2 \cdot 1 + 7 \cdot 1) = 4$$

Example: let $\boldsymbol{x}$ and $\boldsymbol{y}$ be zero mean vectors containing n measurements on two variables $(n^{-1}\boldsymbol{i}'\boldsymbol{x} = n^{-1}\boldsymbol{i}'\boldsymbol{y} = 0)$. Then $C_{xy} = n^{-1}\boldsymbol{x}'\boldsymbol{y}$ is the sample covariance.

Vector Norm or length

By Pythagoras theorem, the norm of $\boldsymbol{a}$ is the square root of the inner product of $\boldsymbol{a}$ with itself:

$$\| \boldsymbol{a} \| = \sqrt{\boldsymbol{a}'\boldsymbol{a}} = \left(\sum_{i=1}^{n} a_i^2 \right)^{1/2}$$

This is the distance from the origin of the point $\boldsymbol{a}$ or the length of the vector.

- $||\boldsymbol{a}|| \geq 0$
- The (normalized) vector $\boldsymbol{a}/||\boldsymbol{a}||$ has unit length.

Example: let $\boldsymbol{x}$ a zero mean vector. Then $SD_x = n^{-1/2}||\boldsymbol{x}||$ is the sample standard deviation of X; $D_x = ||\boldsymbol{x}||^2$ is the deviance of X and $V_x = D_x/n$ is the variance.

Euclidean distance

The distance between the vectors $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is the norm of the difference vector $\boldsymbol{x}_i - \boldsymbol{x}_j$:

$$d_{ij} = ||\boldsymbol{x}_i - \boldsymbol{x}_j|| = \sqrt{(\boldsymbol{x}_i - \boldsymbol{x}_j)'(\boldsymbol{x}_i - \boldsymbol{x}_j)} = \left(\sum_{k=1}^{p}(x_{ik} - x_{jk})^2\right)^{1/2}$$

Orthogonality

Two vectors are orthogonal, $\boldsymbol{a} \perp \boldsymbol{b}$, iff their inner product is zero, $\boldsymbol{a}'\boldsymbol{b} = 0$.

Examples

$$\boldsymbol{a} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad ||\boldsymbol{a}|| = ||\boldsymbol{b}|| = \sqrt{2}, \boldsymbol{a}'\boldsymbol{b} = 0$$

$$\boldsymbol{c} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \boldsymbol{d} = \begin{bmatrix} .6 \\ .2 \end{bmatrix},$$

$$||\boldsymbol{c}|| = \sqrt{10} = 3.16, ||\boldsymbol{d}|| = \sqrt{.4} = 0.63, \boldsymbol{c}'\mathbf{d} = 0$$

$$\boldsymbol{a}'\boldsymbol{c} = -2$$

Cauchy-Schwartz inequality:

$$|\boldsymbol{a}'\boldsymbol{b}| \leq \ ||\boldsymbol{a}|| \cdot ||\boldsymbol{b}||$$

(equality holds iff $\boldsymbol{a} = \rho\boldsymbol{b}, \rho \geq 0$).
This is used to prove that $|C_{xy}| \leq SD_x \cdot SD_y$
Triangular inequality:

$$||\boldsymbol{a} + \boldsymbol{b}|| \leq ||\boldsymbol{a}|| + ||\boldsymbol{b}||$$

(equality holds iff $\boldsymbol{a} \perp \boldsymbol{b}$).

Geometric interpretation

Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{V} \subseteq \mathbb{R}^n$.

- We can choose a scalar ρ and a vector $\boldsymbol{c}$, orthogonal to $\boldsymbol{b}$, $\boldsymbol{b} \perp \boldsymbol{c}$, so that, by the parallelogram law, $\boldsymbol{a} = \boldsymbol{c} + \rho\boldsymbol{b}$.
- The vector $\hat{\boldsymbol{a}} = \rho\boldsymbol{b}$ is the orthogonal projection of $\boldsymbol{a}$ along the vector $\boldsymbol{b}$.
- The inner product of both sides of $\boldsymbol{a} = \boldsymbol{c} + \rho\boldsymbol{b}$ and the vector $\boldsymbol{b}$ yields

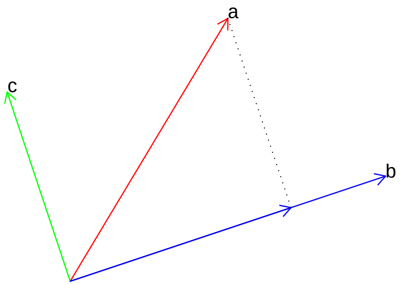
$$\boldsymbol{a}'\boldsymbol{b} = \boldsymbol{c}'\boldsymbol{b} + \rho\boldsymbol{b}'\boldsymbol{b} = \rho||\boldsymbol{b}||^2.$$

- Thus, $\rho = \boldsymbol{a}'\boldsymbol{b}/||\boldsymbol{b}||^2$, and we can write the projection of $\boldsymbol{a}$

$$\hat{\boldsymbol{a}} = \frac{\boldsymbol{a}'\boldsymbol{b}}{||\boldsymbol{b}||^2}\boldsymbol{b}$$

(we can rearrange the formula in this way: $\hat{\boldsymbol{a}} = \boldsymbol{b}(\boldsymbol{b}'\boldsymbol{b})^{-1}\boldsymbol{b}'\boldsymbol{a}$)

Figure: Geometric interpretation of inner product



- Note: $\rho = (\boldsymbol{b}'\boldsymbol{b})^{-1}\boldsymbol{b}'\boldsymbol{a}$ represents the coordinate of the point $\boldsymbol{a}$ in the subspace generated by the vector $\boldsymbol{b}$.
- Denoting by θ the angle formed by the two vectors $\boldsymbol{a}$ e $\boldsymbol{b}$,

$$|\cos\theta| = \frac{||\rho\boldsymbol{b}||}{||\boldsymbol{a}||} = \frac{|\boldsymbol{a}'\boldsymbol{b}|}{||\boldsymbol{a}||\;||\boldsymbol{b}||}$$

or

$$\boldsymbol{a}'\boldsymbol{b} = ||\boldsymbol{a}||\;||\boldsymbol{b}||\cos\theta.$$

- Note: $\boldsymbol{a} \perp \boldsymbol{b} \Rightarrow \theta = \frac{\pi}{2} + k\pi, k = 0, \pm 1, \pm 2, \ldots$.

Example: regression and correlation

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be zero mean vectors containing n measurements on two variables X and Y.
The scalar

$$\frac{\boldsymbol{x}'\boldsymbol{y}}{||\boldsymbol{x}||^2} = (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{y} = \frac{C_{xy}}{V_x}$$

is the coefficient of the regression of Y on X and

$$\hat{\boldsymbol{y}} = \frac{C_{xy}}{V_x}\boldsymbol{x}$$

is the vector of predicted values of Y.

$$\frac{\sum_i x_i y_i}{\sqrt{(\sum_i x_i^2)(\sum_i y_i^2)}}$$

is the sample correlation coefficient.

Matrices

A matrix is a rectangular ($n \times p$) or two-dimensional array of numbers.

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2k} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \dots & x_{ik} & \dots & x_{ip} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} & \dots & x_{np} \end{bmatrix}.$$

The rows or columns can be considered as vectors.
We can also write $\boldsymbol{X} = \{x_{ik}\}$. In a typical data matrix, the index $i = 1, \dots, n$, refers to the statistical units, and the index $k = 1, \dots, p$, to the variables or attributes.

We can represent $\boldsymbol{X}$ as a partitioned matrix whose generic block is the $1 \times p$ row vector $\boldsymbol{x}_i' = [x_{i1}, x_{i2}, \ldots, x_{ik}, \ldots, x_{ip}]$, which contains the profile of the i-th row unit.

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1' \\ \vdots \\ \boldsymbol{x}_i' \\ \vdots \\ \boldsymbol{x}_n' \end{bmatrix}.$$

Alternatively, we can partition

$$\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, \ldots, \boldsymbol{x}_p],$$

where $\boldsymbol{x}_k$ is the $n \times 1$ column vector referring to the k-th variable.

Matrix manipulations

Matrix transpose: transposition yields the $m \times n$ matrix with rows and columns interchanged:

$$\boldsymbol{X}' = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{i1} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{i2} & \dots & x_{n2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1k} & x_{2k} & \dots & x_{ik} & \dots & x_{nk} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1p} & x_{2p} & \dots & x_{ip} & \dots & x_{np} \end{bmatrix},$$

Scalar multiplication: $\rho\boldsymbol{X} = \{\rho x_{ik}\}$,
Matrix sum: $\boldsymbol{C} = \boldsymbol{X} + \boldsymbol{Y}, c_{ik} = \{x_{ik} + y_{ik}\}$

Matrix product

let $\boldsymbol{A}$ be an e $n \times m$ matrix whose i-th row is the $1 \times m$ vector $\boldsymbol{a}_i'$. Let also $\boldsymbol{B}$ be an $m \times p$ matrix whose j-th column is the $m \times 1$ vector $\boldsymbol{b}_j$, so that

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1' \\ \vdots \\ \boldsymbol{a}_i' \\ \vdots \\ \boldsymbol{a}_n' \end{bmatrix}, \qquad \boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_j, \ldots, \boldsymbol{b}_p]$$

The matrix product $\boldsymbol{C} = \boldsymbol{AB}$, where $\boldsymbol{A}$ pre-multiplies $\boldsymbol{B}$, is the $n \times p$ matrix with elements

$$c_{ij} = \boldsymbol{a}_i' \boldsymbol{b}_j = \sum_{k=1}^{m} a_{ik} b_{kj}, i = 1, \ldots, n; j = 1, \ldots, p.$$

Properties

- $(\boldsymbol{A}')' = \boldsymbol{A}$
- $(\boldsymbol{A} + \boldsymbol{B})' = \boldsymbol{A}' + \boldsymbol{B}'$
- $(\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}'$.
- $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$
- $\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}$

Il matrix product is not commutative: if $n \neq p$, $\boldsymbol{BA}$ is not even defined.
Notice the difference between $\boldsymbol{X}'\boldsymbol{X}$ ($p \times p$ matrix of crossproducts) and $\boldsymbol{XX}'$ ($n \times n$)

Examples:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 5 & -1 \\ 3 & 2 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 2 & -1 \\ 3 & 6 \end{bmatrix}, \boldsymbol{i} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\boldsymbol{C} = \boldsymbol{AB} = \begin{bmatrix} 2 & -1 \\ 7 & -11 \\ 12 & 9 \end{bmatrix}, \boldsymbol{i}'\boldsymbol{C} = [21 \quad -3]$$

$\boldsymbol{BA}$ is not defined.

Mean vector:

$$\frac{1}{n}\boldsymbol{X}'\boldsymbol{i} = \bar{\boldsymbol{x}} \qquad \bar{\boldsymbol{x}} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_k \\ \vdots \\ \bar{x}_p \end{bmatrix}, \quad (p \times 1)$$

$$\bar{\boldsymbol{x}}' = \frac{1}{n}\boldsymbol{i}'\boldsymbol{X} = [\bar{x}_1, \dots, \bar{x}_k, \dots, \bar{x}_p], \quad (1 \times p)$$

Important special cases

- A square matrix has many rows as columns ($m = n$).
- A square matrix $\boldsymbol{A}$ is *symmetric* if $\boldsymbol{A}' = \boldsymbol{A}$.
- *Diagonal matrix*: a square matrix with nonzero diagonal entries

$$\boldsymbol{D} = \begin{bmatrix} d_1 & 0 & \dots & 0 & 0 \\ 0 & d_2 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & d_{n-1} & 0 \\ 0 & 0 & \dots & 0 & d_n \end{bmatrix} = \text{diag}(d_1, \dots, d_n)$$

- Identity matrix: if $\boldsymbol{A}$ is $n \times n$, $\boldsymbol{I}_n\boldsymbol{A} = \boldsymbol{A}$, $\boldsymbol{A}\boldsymbol{I}_m = \boldsymbol{A}$.

$$\boldsymbol{I} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix},$$

- *Scalar matrix*: $\rho \boldsymbol{I}$
- Quadratic form: let $\boldsymbol{A}$ be an n dimensional square matrix and $\boldsymbol{x}$ an $n \times 1$ vector. The scalar $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ is called a quadratic form.
- A symmetric matrix $\boldsymbol{A}$ such that, for any vector $\boldsymbol{x}$, $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$ is said to be semi-positive (nonnegative) definite. If $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0, \forall \boldsymbol{x} \in \mathbb{R}^n$, the matrix is positive definite (p.d.).
- Outer product: if $\boldsymbol{x}$ an $n \times 1$ vector, the outer product $\boldsymbol{x}\boldsymbol{x}'$ is an $n \times n$ matrix.

Example: variance covariance matrix $\boldsymbol{S}$ $(p \times p)$:

$$\boldsymbol{S} = \begin{bmatrix} s_1^2 & s_{12} & \dots & s_{1k} & \dots & s_{1p} \\ s_{21} & s_2^2 & \dots & s_{2k} & \dots & s_{2p} \\ \vdots & \vdots & \ddots & \dots & \vdots & \vdots \\ s_{k1} & s_{k2} & \dots & s_k^2 & \dots & s_{kp} \\ \vdots & \vdots & \dots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \dots & s_{pk} & \dots & s_p^2 \end{bmatrix},$$

$$s_k^2 = \frac{1}{n}\sum_{i=1}^{n}(x_{ik} - \bar{x}_k)^2, \; s_{hk} = \frac{1}{n}\sum_{i=1}^{n}(x_{ih} - \bar{x}_h)(x_{ik} - \bar{x}_k)$$

$\boldsymbol{S}$ is symmetric $(s_{hk} = s_{kh})$, and p.d.

In terms of the original data matrix $\boldsymbol{X}$,

$$\boldsymbol{S} = \frac{1}{n}(\boldsymbol{X} - \boldsymbol{i}_n\bar{\boldsymbol{x}}')'(\boldsymbol{X} - \boldsymbol{i}_n\bar{\boldsymbol{x}}') = \frac{1}{n}\boldsymbol{X}'\boldsymbol{X} - \bar{\boldsymbol{x}}\bar{\boldsymbol{x}}',$$

where $\boldsymbol{i}_n$ is an $n \times 1$ of ones.

In terms of the unit profiles

$$\boldsymbol{S} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})'$$

Matrix rank

Let $\boldsymbol{A}$ be $n \times m$.

- Column (row) space: the vector space generated by the column (row) vectors that form the matrix $\boldsymbol{A}$.
- Column (row) rank: the dimension of the vector space generated by the columns (rows) of the matrix $\boldsymbol{A}$.
- The column and row rank are coincident and so we can define the rank of the matrix as the maximum number of linearly independent vectors (those forming either the rows or the columns) and denote it by $\mathrm{r}(\boldsymbol{A})$. Obviously, $\mathrm{r}(\boldsymbol{A}) \leq \min(n, m)$.
- Assume $m \leq n$. If $\mathrm{r}(\boldsymbol{A}) = m$ the matrix is full column rank. If further $m = n$, $\boldsymbol{A}$ is full rank.
- Note: if $\boldsymbol{Ab} = \mathbf{0}$ for $\boldsymbol{b} \neq \mathbf{0}$, the columns of $\boldsymbol{A}$ are linearly dependent and $\boldsymbol{A}$ is reduced rank.

Determinant

Let $\boldsymbol{A}$ be $n \times n$.

- Its determinant, $\det(\boldsymbol{A})$, or $|\boldsymbol{A}|$, is a scalar, whose absolute value measures the volume of the parallelogram delimited in $\mathbb{R}^n$ by the columns of $\boldsymbol{A}$.
- For the identity matrix, $|\boldsymbol{I}| = 1$.
- For $\boldsymbol{D} = \text{diag}(d_1, \ldots, d_n)$

$$|\boldsymbol{D}| = d_1 \cdot d_2 \cdots d_n = \prod_{i=1}^{n} d_i$$

- Moreover, if ρ is a scalar $|\rho \boldsymbol{D}| = \rho^n |\boldsymbol{D}|$.
- If $\boldsymbol{A}$ is 2×2,

$$|\boldsymbol{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

The general expression for the determinant is the following Laplace (cofactor) expansion:

$$|\boldsymbol{A}| = \sum_{j=1}^{n} a_{ij}(-1)^{i+j}|\boldsymbol{A}_{ij}|,$$

where $\boldsymbol{A}_{ij}$ is the submatrix obtained from $\boldsymbol{A}$ by removing the i-th row and the j-th column; $|\boldsymbol{A}_{ij}|$ is called a minor of $\boldsymbol{A}$ and $(-1)^{i+j}|\boldsymbol{A}_{ij}|$ is a cofactor.
(We do not need to bother about this formula. The computation of the determinant is usually carried out by transforming the matrix into a simpler form, e.g. triangular).
Properties

- If the columns (rows) of $\boldsymbol{A}$ are linearly dependent, so that $\text{rank}(\boldsymbol{A}) < n$, $|\boldsymbol{A}| = 0$.
- $|\boldsymbol{AB}| = |\boldsymbol{A}|\,|\boldsymbol{B}|$.
- $|\boldsymbol{A}'| = |\boldsymbol{A}|$.

Trace of a matrix

The trace of a square matrix is the sum of its diagonal elements.
If $\boldsymbol{A}$ is $n \times n$,

$$\operatorname{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} a_{ii}$$

Properties:

- $\operatorname{tr}(\rho\boldsymbol{A}) = \rho\operatorname{tr}(\boldsymbol{A})$
- $\operatorname{tr}(\boldsymbol{A}+\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{A}) + \operatorname{tr}(\boldsymbol{B})$
- $\operatorname{tr}(\boldsymbol{A}') = \operatorname{tr}(\boldsymbol{A})$
- $\operatorname{tr}(\boldsymbol{AB}) = \operatorname{tr}(\boldsymbol{BA})$, if both products exist.

Linear equations systems

Consider the system of n linear equations in n unknowns, where $\boldsymbol{A}$ is a known $n \times n$ coefficients matrix and $\boldsymbol{b}$ a known $n \times 1$ vector:

$$\boldsymbol{Ax} = \boldsymbol{b}.$$

- The system is said to be homogeneous if $\boldsymbol{b} = \mathbf{0}$.
- In the latter case there exists non trivial solutions ($\boldsymbol{x} \neq \mathbf{0}$) iff $\boldsymbol{A}$ has reduced rank ($\text{rank}(\boldsymbol{A}) < n$) or the columns of $\boldsymbol{A}$ are linearly dependent, or equivalently $|\boldsymbol{A}| = 0$.
- A non homogeneous system admits a unique solution iff $|\boldsymbol{A}| \neq 0$, or equivalently $\text{rank}(\boldsymbol{A}) = n$. In such case the solution can be written is

$$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$$

where $\boldsymbol{A}^{-1}$ is the inverse of $\boldsymbol{A}$.

Matrix Inverse

Let $\boldsymbol{A}$ be a square matrix of dimension p with full rank ($\mathrm{r}(A) = p$).

- The inverse matrix is the matrix $\boldsymbol{B}$ which when pre-multiplied or post-multiplied by $\boldsymbol{A}$ returns the identity matrix:

$$\boldsymbol{BA} = \boldsymbol{I}, \quad \boldsymbol{AB} = \boldsymbol{I}.$$

  In the sequel we shall write $\boldsymbol{B} = \boldsymbol{A}^{-1}$.
- The inverse exists and is unique iff $\mathrm{r}(A) = p$, in which case we say that $\boldsymbol{A}$ is non singular or invertible.
- For a diagonal matrix the computation of the inverse is immediate: $\boldsymbol{D}^{-1} = \mathrm{diag}(1/d_1, \ldots, 1/d_n)$.
- In general, the computation of the inverse entails the solution of a system of p^2 linear equations in p^2 unknowns.

The solution can be expressed as follows

$$\boldsymbol{A}^{-1} = \frac{1}{|\boldsymbol{A}|}\boldsymbol{A}^*$$

where $\boldsymbol{A}^*$ is known as the *adjoint matrix* of $\boldsymbol{A}$, with elements given by the cofactors of $\boldsymbol{A}$:

$$a^*_{ji} = (-1)^{i+j}|\boldsymbol{A}_{ij}|$$

We now illustrate the 2×2 case. Form the definition of an inverse, $\boldsymbol{AB} = \boldsymbol{I}$, it follows

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

This yields a system of 4 equations in 4 unknowns with solution:

$$\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}.$$

Here, we have

$$|\boldsymbol{A}| = a_{11}a_{22} - a_{12}a_{21}, \quad \boldsymbol{A}^* = \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}.$$

Examples:

$$\boldsymbol{S} = \text{diag}\{10, 20, 1/2\}; \quad \boldsymbol{S}^{-1} = \text{diag}\{0.1, 0.05, 2\}$$

$$\boldsymbol{S} = \text{diag}\{10, 20, 0\}; \quad \boldsymbol{S}^{-1} \text{ does not exist}$$

$$\boldsymbol{R} = \begin{bmatrix} 1.0 & 0.2 \\ 0.2 & 1.0 \end{bmatrix}; \boldsymbol{R}^{-1} = \begin{bmatrix} 1.04 & -0.21 \\ -0.21 & 1.04 \end{bmatrix}$$

$$\boldsymbol{R} = \begin{bmatrix} 1.0 & 0.9 \\ 0.9 & 1.0 \end{bmatrix}; \boldsymbol{R}^{-1} = \begin{bmatrix} 5.26 & -4.74 \\ -4.74 & 5.26 \end{bmatrix}$$

$$\boldsymbol{R} = \begin{bmatrix} 1.0 & -0.9 \\ -0.9 & 1.0 \end{bmatrix}; \boldsymbol{R}^{-1} = \begin{bmatrix} 5.26 & 4.74 \\ 4.74 & 5.26 \end{bmatrix}$$

$$\boldsymbol{R} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}; \boldsymbol{R}^{-1} \text{ does not exist}$$

$$\boldsymbol{R} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}; \boldsymbol{R}^{-1} \text{ does not exist}$$

$$\boldsymbol{X} = \begin{bmatrix} 10 & 20 & 15 \\ 15 & 15 & 15 \\ 5 & 3 & 4 \\ 20 & 32 & 26 \end{bmatrix}; (\boldsymbol{X}'\boldsymbol{X})^{-1} \text{ does not exist}$$

Properties of the matrix inverse

Hereby we list some useful properties of matrix inverse.

- $|\boldsymbol{A}^{-1}| = 1/|\boldsymbol{A}|$
- $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$
- $(\boldsymbol{A}^{-1})' = (\boldsymbol{A}')^{-1}$
- $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$

Orthogonal matrix A square matrix is orthogonal if the transpose and the inverse coincide

$$\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I} \quad \boldsymbol{A}\boldsymbol{A}' = \boldsymbol{I}$$

Vector Spaces and Bases

Let $\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1 \cdot \boldsymbol{e}_1 + 1 \cdot \boldsymbol{e}_2 = \boldsymbol{i}; \quad \begin{bmatrix} 0.7 \\ -0.3 \end{bmatrix} = 0.7 \cdot \boldsymbol{e}_1 - 0.3 \cdot \boldsymbol{e}_2$$

All the points lying on the plane $[x_1, x_2]' \in \mathbb{R}^2$ can be generated by linearly combining $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$.
The set of vectors in $\mathbb{R}^2$ will be call a vector space, while $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ form a basis of that vector space.

- A vector space is a set of vectors that is closed with respect to scalar multiplication and vector sum ($\boldsymbol{a}, \boldsymbol{b} \in \mathbb{V} \Rightarrow \boldsymbol{c} = \rho_1 \boldsymbol{a} + \rho_2 \boldsymbol{b} \in \mathbb{V}$, where $\mathbb{V}$ is a VS).
- Any set of n-vectors such that any linear combination of the vectors in $\mathbb{V} \subseteq \mathbb{R}$ is also in $\mathbb{V}$ constitutes a basis of the vector space.

Example i. $\boldsymbol{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} 3 \\ .5 \end{bmatrix}$
form a basis for $\mathbb{R}^2$. Note: the basis is not unique.
Example ii. $\boldsymbol{a} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$
do not form a basis for $\mathbb{R}^2$, but only for $\mathbb{R}$ (the vector $[c_1, c_2]'$, $c_2 \neq 0$, belonging to $\mathbb{R}^2$, cannot be obtain from a linear combination $\boldsymbol{a}$ and $\boldsymbol{b}$).

Note: $\boldsymbol{a}$ e $\boldsymbol{b} = \rho\boldsymbol{a}$ are not a basis for $\mathbb{R}^2$.

Linear dependence

- If a given vector can be formed by a linear combination of one or more vectors, the set of vectors (including the given one) is said to be linearly dependent; conversely, if in a set of vectors no one vector can be represented as a linear combination of any of the others, the set of vectors is said to be linearly independent.
- A set of vectors is linearly dependent if one or more of its elements can be formed by a linear combination of any of the others elements of the set, or, equivalently, there exists a non trivial (i.e. with nonzero coefficients) linear combination yielding the null vector.
- For instance, the three vectors

$$\boldsymbol{a} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{c} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

are linearly dependent as $\boldsymbol{c} = 2\boldsymbol{a} - \boldsymbol{b}$, or equivalently, $2\boldsymbol{a} - \boldsymbol{b} - \boldsymbol{c} = \mathbf{0}$.

Linear independence

A set of vectors $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$ is linearly independent if

$$\rho_1 \boldsymbol{a}_1 + \rho_2 \boldsymbol{a}_2 + \cdots + \rho_n \boldsymbol{a}_n = \mathbf{0} \iff \rho_1 = \cdots = \rho_n = 0$$

The basis of an n-dimensional vector space is formed by any set of n linearly independent vectors.
Example:

$$\boldsymbol{a} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{c} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$$

while $\{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\}$ are linearly dependent ($2\boldsymbol{a} - \boldsymbol{b} - \boldsymbol{c} = \mathbf{0}$).
$\boldsymbol{a}, \boldsymbol{b}, \mathbf{u}$ are linearly independent.

Dimension of a vector space

- The dimension of a vector space $\mathbb{V} \subseteq \mathbb{R}^n$ generated by n vectors is the largest number ($\leq n$) of linearly independent vectors that generate $\mathbb{V}$.
- Example: the dimension of the vector space generated by $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c} = \rho_1 \boldsymbol{a} + \rho_2 \boldsymbol{b}$, with $\boldsymbol{a}$ and $\boldsymbol{b}$ linearly independent is 2.