Special Chapters on Artificial Intelligence

Lecture 3. Matrix Algebra and Statistics

Cristian Gatu

[1]Faculty of Computer Science
"Alexandru Ioan Cuza" University of Iași, Romania

MCO, MDS, 2018–2019

In modelling, a lot of problems are linear, or approximated by linear models. Such problems are solved by MATRIX METHODS.

Contents

Content

Variance

- The objective is to account for, or explain, the variation in the data.
- Variance is the most commonly used measure of dispersion in the data.
- Variance directly proportional to the amount of variation or information in the data.

The data below gives two financial ratios, X_1 and X_2, for 12 hypothetical companies.

	Original Data		Mean-Corrected Data		Standardize Data	
Firm	X_1	X_2	x_1	x_2	x_1	x_2
1	13	4	7.92	3.83	1.62	1.11
2	10	6	4.92	5.83	1.01	1.69
3	10	2	4.92	1.83	1.01	0.53
4	8	-2	2.92	-2.17	0.60	-0.63
5	7	4	1.92	3.83	0.39	1.11
6	6	-3	0.92	-3.17	0.19	-0.92
7	5	0	-0.08	-0.17	-0.02	-0.05
8	4	2	-1.08	1.83	-0.22	0.53
9	2	-1	-3.08	-1.17	-0.63	-0.34
10	0	-5	-5.08	-5.17	-1.04	-1.49
11	-1	-1	-6.08	-1.17	-1.24	-0.34
12	-3	-4	-8.08	-4.17	-1.65	-1.20
Mean	5.08	0.17	0	0	0	0
SS			262.92	131.67	11	11
Var	23.90	11.97	23.90	11.97	1	1

Mean. Variance

- The MEAN of the jth variable:

$$\boxed{\mu_j = \frac{\sum_{i=1}^{n} X_{ij}}{n}}$$

  where X_{ij} is the ith observation of the jth variable and n is the number of observations.

- The MEAN-CORRECTED jth variable is $x_{ij} = X_{ij} - \mu_j$.
- The VARIANCE of the jth variable:

$$\boxed{s_{jj} = \frac{\sum_{i=1}^{n} x_{ij}^2}{n-1} = \frac{\mathrm{SS}}{\mathrm{df}}}$$

  where SS is the *sum of squares* deviations from the mean and df is the degree of freedom.

Covariance

- COVARIATION describes the linear relationship, or association, between two variables
- COVARIANCE is a measure of the covariation between two variables X_i and X_j:

$$\boxed{s_{ij} = \frac{\sum_{k=1}^{n} x_{ki} x_{kj}}{n-1} = \frac{\text{SCP}}{\text{df}}}$$

where SCP is the *Sum of the Cross Products* (SCP).

Sum of Squares and Cross Products

- The SS and SCP are summarized in a SUM OF SQUARES AND CROSS PRODUCTS (**SSCP**) matrix.
- The variance and covariances are usually summarized in a *covariance* **S** matrix.
- The **SSCP** and **S** of the two financial ratios are given by:

$$\mathbf{SSCP} = \begin{pmatrix} 262.92 & 136.38 \\ 136.38 & 131.67 \end{pmatrix} \quad \text{and} \quad \mathbf{S} = \begin{pmatrix} 23.90 & 12.40 \\ 12.40 & 11.97 \end{pmatrix}.$$

Note that the matrices are symmetric.

Variance. Covariance

- The variance of a given variable is a measure of its variation in the data. The variances of variables can only be compared if the variables are measured using the same units.

- The Covariance between two variables is a measure of covariation between them. The absolute value of the lower bound covariance is zero implying that the two variables are not linearly associated. However it has no upper bound and this makes it difficult to compare the association between two variables across data sets.

Standardization

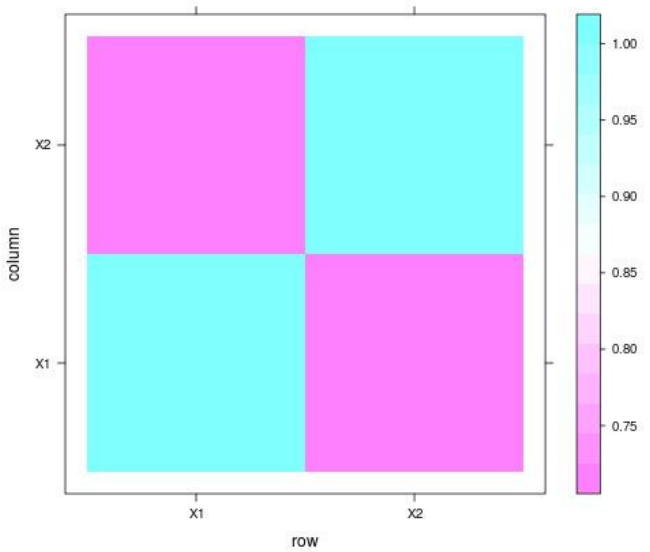
- Standardized data are obtained by dividing the mean-corrected data by the respective standard deviation (square root of variance).
- The variance of the standardized variables is always 1.
- The covariation of standardize variables are always lie between -1 and 1. The value will be:
  - 0 (zero) : no linear relationship between the two variables;
  - -1 (minus one) : a perfect inverse linear relationship;
  - $+1$ (plus one) : a perfect direct linear relationship.

Correlation matrix

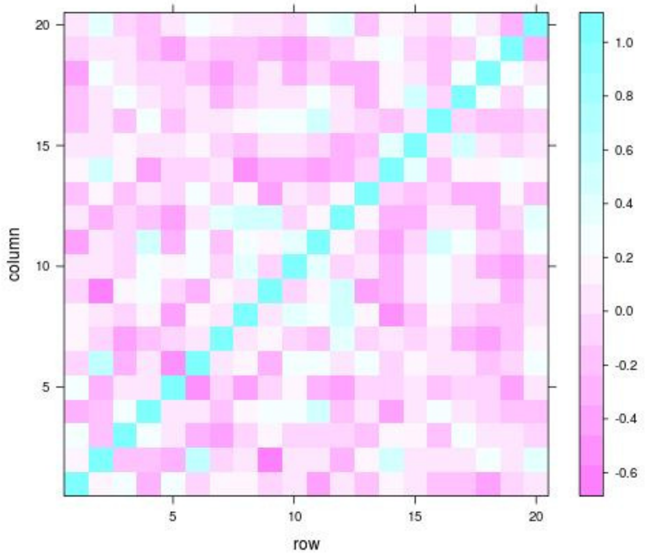
- The covariance of two standardized variables is called the CORRELATION COEFFICIENT.
- The CORRELATION MATRIX ($\mathbf{R}$) is the covariance matrix for standardized data.
- In the example the correlation matrix is:

$$\mathbf{R} = \begin{pmatrix} 1.00 & 0.733 \\ 0.733 & 1.00 \end{pmatrix}.$$

Correlation matrix for the two ratio example



Correlation matrix of 20 variables



Content

Matrices

- An $m \times n$ MATRIX A containing $m \times n$ elements has form:

$$A = \left(\begin{array}{cccccc} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mj} & \cdots & a_{mn} \end{array}\right) \begin{array}{l} \\ \\ \\ \leftarrow i\text{th row} \\ \\ \\ \end{array}$$

- The subscripts of an element a_{ij} indicates that the element is located at the interception of row i and column j, where $1 \leq i \leq m$ and $1 \leq j \leq n$.

Matrices

- A matrix with one row or one column are called ROW VECTORS or COLUMN VECTORS, respectively.
- A row vector R having n real elements is denoted by $R \in \Re^{1\times n}$ and has the general form $R = (r_1 \; \dots \; r_n)$.
- A column vector C having m real elements is denoted by $C \in \Re^{m\times 1}$ and has the general form

$$C = \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}.$$

- Generally a C m-elements real vector will be assumed to be a column vector and denoted by $C \in \Re^m$.

Special types of matrices

- SQUARE MATRIX: an $m \times n$ matrix is square if $m = n$.
- IDENTITY (OR UNIT) MATRIX: I_m.
- TRANSPOSED OF A MATRIX: if $A = [a_{ij}] \in \Re^{m \times n}$, $B = [b_{ij}] \in \Re^{n \times m}$ and $b_{ji} = a_{ij}$ then $B = A^T$.
- SYMMETRIC MATRIX: $A = A^T$.
- UPPER TRIANGULAR MATRIX: $U = [a_{ij}] \in \Re^{m \times n}$ s.t. $\forall i > j$, $u_{ij} = 0$
- LOWER TRIANGULAR MATRIX: $L = [a_{ij}] \in \Re^{m \times n}$ s.t. $\forall i - j < m - n$, $l_{ij} = 0$.

Matrix operations

- Two matrices can be ADDED or SUBTRACTED (element by element) iff they have the same dimension.
- The MULTIPLICATION OF A SCALAR BY A MATRIX is equivalent into multiplying each element of the matrix by the scalar.
- The INNER PRODUCT is an operation between a row and a column vector (in this order). It is computed by multiplying corresponding elements in the two vectors and algebraically summing.
- MATRIX MULTIPLICATION. Given $A \in \Re^{m_a \times n_a}$ and $B \in \Re^{m_b \times n_b}$ the matrix product $C = A\,B$ is defined iff $n_a = m_b$. The element c_{ij} is defined to be the inner product of row i in matrix A and column j in matrix B.

Partitioned matrices

- A partitioned matrix contains sub-matrices as elements.
- E.g. consider the partitioning of $A, B \in \Re^{m\times n}$ as:

$$A = \begin{array}{cc} \begin{array}{ccc} n_1 & \quad & n_N \end{array} & \\ \begin{pmatrix} A_{11} & \dots & A_{1N} \\ \vdots & & \vdots \\ A_{M1} & \dots & A_{MN} \end{pmatrix} & \begin{array}{c} m_1 \\ \\ m_M \end{array} \end{array} \text{ and } B = \begin{array}{cc} \begin{array}{ccc} n_1 & \quad & n_N \end{array} & \\ \begin{pmatrix} B_{11} & \dots & B_{1N} \\ \vdots & & \vdots \\ B_{M1} & \dots & B_{MN} \end{pmatrix} & \begin{array}{c} m_1 \\ \\ m_M \end{array} \end{array},$$

where $n = \sum_{i=1}^{N} n_i$ and $m = \sum_{i=1}^{M} m_i$.

- addition and multiplication of partitioned matrices.

$$A + B = \begin{pmatrix} A_{11} + B_{11} & \dots & A_{1N} + B_{MN} \\ \vdots & & \vdots \\ A_{M1} + B_{M1} & \dots & A_{MN} + B_{MN} \end{pmatrix}.$$

Rank of a matrix

- The number of linearly independent columns of a matrix is called COLUMN RANK, hereafter RANK. It will be denoted by $\text{rank}(A)$.
- The square matrix $A \in \Re^{n\times n}$ is said to be *non-singular* if the $\text{rank}(A) = n$. Otherwise it is called *singular*.
- Properties
  1. $\text{rank}(A) = \text{rank}(A^T)$.
  2. $\text{rank}(A) = \text{rank}(A^T A) = \text{rank}(AA^T)$.
  3. The rank of A is unchanged by pre- or postmultiplication of A by a non-singular matrix.

Trace of a matrix

- For a square matrix $A = [a_{ij}] \in \Re^{n\times n}$ the sum of its diagonal elements is called its trace, i.e.

$$\boxed{\text{trace}(A) = \sum_{i=1}^{n} a_{ii}}.$$

- Properties

1. $\text{trace}(A) = \text{trace}(A^T)$.
2. $\text{trace}(AB) = \text{trace}(BA)$.
3. $\text{trace}(ABC) = \text{trace}(BCA) = \text{trace}(CAB)$.
4. $\text{trace}(A+B) = \text{trace}(B+A) = \text{trace}(A) + \text{trace}(B)$.
5. $\text{trace}\left(\sum_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} \text{trace}(A_i)$.
6. $\text{trace}(\kappa A) = \kappa\, \text{trace}(A)$.

Matrix properties

- For any two matrices A and B, it CANNOT be stated that $AB = BA$.
- If A is an $m \times n$ matrix, then $I_m A = A I_n = A$.
- $(AB)^T = B^T A^T$. Generally: $(A_1 A_2 \cdots A_n)^T = A_n^T \cdots A_2^T A_1^T$.

Inverse of a matrix

- The relationship between a square matrix A and its inverse, denoted by A^{-1} (inverse of A), is that:

$$\boxed{A^{-1}A = AA^{-1} = I}.$$

- Note that
  - The matrix A must be square.
  - The dimensions of A and A^{-1} are the same.
  - Only non-singular matrices have an inverse.
- For $|A| \neq 0$, the inverse of A is given by:

$$\boxed{A^{-1} = \frac{1}{|A|} A_C^T}.$$

Gaussian reduction procedure

- Consider the $m \times m$ matrix A. Construct the augmented matrix $(A \mid I_m)$.

- The Gaussian elimination method transforms $(A \mid I_m)$ to $(I_m \mid A^{-1})$ by applying two basic operations:
  1. Rows can be multiplied by a non zero constant; and
  2. non zero multiples of one row can be added to another row.

Properties of the inverse

- The inverse of a symmetric matrix is also symmetric.
- $(A^T)^{-1} = (A^{-1})^T = A^{-T}$.
- Let $A_1, \ldots, A_n \in \Re^{n \times n}$. Then, $(A_1 A_2 \cdots A_n)^{-1} = (A_n^{-1} \cdots A_2^{-1} A_1^{-1})$.
- If c is a non zero scalar, then $(cA)^{-1} = \frac{1}{c} A^{-1}$.
- The inverse of a diagonal matrix is a diagonal matrix consisting of the reciprocals of the original elements.
- The inverse of a triangular matrix is also triangular.

System of equations

Consider the $n \times n$ system of equations having the form:

$$\begin{array}{llll} a_{11}x_1 & +a_{12}x_2 & +\dots & +a_{1n}x_n & = b_1 \\ a_{21}x_1 & +a_{22}x_2 & +\dots & +a_{2n}x_n & = b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{n1}x_1 & +a_{n2}x_2 & +\dots & +a_{nn}x_n & = b_n \end{array}$$

can be written in a matrix form as

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \tag{1}$$

$$\text{or} \quad Ax = b \tag{2}$$

System of equations

- Assume that the equations are linear independent, that is, A is not singular (it has inverse).

- Premultiply both sides of (2) by A^{-1} it gives:

$$A^{-1}Ax = A^{-1}b \quad \text{or} \quad x = A^{-1}b$$

  since $A^{-1}A\,x = I_n\,x = x$.

- Thus, the solution of (1) is given by $\boxed{x = A^{-1}b}$.

Orthogonal matrices

- A square matrix $Q \in \Re^{m \times m}$ is orthogonal iff

$$\boxed{Q^T Q = QQ^T = I_m}.$$

- Notice that the inverse of Q is given by Q^T.

- Examples of orthogonal matrices:

$$I_m, \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}.$$

Orthogonal matrices. Property

It preserves the norm (inner product) of a vector.
That is, If $z = Qx$ and Q is orthogonal, then $z^T z = x^T x$.

Note $z^T z = (Qx)^T (Qx) = x^T Q^T Qx = x^T Ix = x^T x$.

Example

$$x = \begin{pmatrix} -1 \\ 3 \end{pmatrix} \quad \text{and} \quad Q = \begin{pmatrix} 0.5 & 0.866 \\ -0.866 & 0.5 \end{pmatrix}.$$

$$z = Qx = \begin{pmatrix} 2.098 \\ 2.366 \end{pmatrix} \quad \text{and} \quad x^T x = 10 = z^T z.$$

Content

Cholesky Decomposition

The CHOLESKY DECOMPOSITION of a symmetric positive definite $n \times n$ matrix A, is given by

$$\boxed{A = LL^T},$$

where $L \in \Re^{n \times n}$ is lower triangular and non-singular. E.g.

$$\begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{12} & A_{22} & A_{23} \\ A_{13} & A_{23} & A_{33} \end{pmatrix} = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix}.$$

Cholesky Decomposition. Example

Let $\quad A = \begin{pmatrix} 5 & 2 & 3 \\ 2 & 10 & 1 \\ 3 & 1 & 2 \end{pmatrix}$

The Cholesky Decomposition of $A = LL^T$ is given by:

$$\begin{pmatrix} 2.24 & 0 & 0 \\ 0.89 & 3.03 & 0 \\ 1.34 & -0.07 & 0.44 \end{pmatrix} \begin{pmatrix} 2.24 & 0.89 & 1.34 \\ 0 & 3.03 & -0.07 \\ 0 & 0 & 0.44 \end{pmatrix}$$

Cholesky Decomposition. Application

Solve the matrix problem $Ax = b$, where A is symmetric and has Cholesky decomposition $A = LL^T$.

Notice that $L(L^T x) = b$ is equivalent to $Lz = b$, where $L^T x = z$. That is, the solution of $Ax = b$ comes in three steps:

1. Compute the Cholesky decomposition $A = LL^T$.
2. Solve the lower-triangular system $Lz = b$ for z.
3. Solve the upper-triangular system $L^T x = z$ for x.

Cholesky Decomposition. Application

Example

Solve $Ax = b$, where $A = \begin{pmatrix} 5 & 2 & 3 \\ 2 & 10 & 1 \\ 3 & 1 & 2 \end{pmatrix}$ and $b = \begin{pmatrix} 7 \\ -16 \\ 5 \end{pmatrix}$.

1. $A = LL^T$, where $L = \begin{pmatrix} 2.24 & 0 & 0 \\ 0.89 & 3.03 & 0 \\ 1.34 & -0.07 & 0.44 \end{pmatrix}$
2. Solve $Lz = b$ which gives $z = \begin{pmatrix} 3.13 \\ 6.19 \\ 0.88 \end{pmatrix}$
3. Solve $L^T x = z$ which gives $x = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}$.

QR decomposition

Let $A \in \Re^{m \times n}$ $(m \geq n)$ have full column rank.

The QR DECOMPOSITION of A has the form:

$$\boxed{A = QR},$$

where $R \in \Re^{n \times n}$ is upper triangular and $Q \in \Re^{m \times m}$ is orthogonal.

QR decomposition. Example

Let $A = Q\begin{pmatrix} R \\ 0 \end{pmatrix}$, where $A \in \Re^{5\times 3}$, $Q \in \Re^{5\times 5}$ is orthogonal and $R \in \Re^{3\times 3}$ is upper-triangular.

$$A = \begin{pmatrix} -8 & -2 & 8 \\ -9 & 7 & 3 \\ -13 & -14 & 17 \\ 4 & 3 & -13 \\ -4 & 1 & 16 \end{pmatrix}, \quad R = \begin{pmatrix} 18.6 & 7.69 & -23.00 \\ 0 & -14.14 & 5.60 \\ 0 & 0 & 15.04 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ and}$$

$$Q = \begin{pmatrix} -0.43 & -0.09 & -0.09 & 0.55 & -0.70 \\ -0.48 & -0.76 & -0.26 & -0.28 & 0.20 \\ -0.70 & 0.61 & -0.17 & -0.05 & 0.33 \\ 0.22 & -0.10 & -0.50 & 0.68 & 0.48 \\ -0.22 & -0.19 & 0.80 & 0.38 & 0.35 \end{pmatrix}.$$

$$Q^TQ = QQ^T = I_5 \quad \text{and} \quad A = QR.$$

QR decomposition. Application

Solve the matrix problem $Ax = b$ using the QR decomposition, where $A \in \Re^{n \times n}$ is non singular.

Let $A = QR$.
The system $Ax = b$ can be written as $QRx = b$.
Premultiply both sides of the system by Q^T it gives:

$$Q^T QRx = Q^T b$$

Since $Q^T Q = I_n$ the latter is equivalent to

$$Rx = Q^T b.$$

QR decomposition. Example

$$A = \begin{pmatrix} 5 & 2 & 3 \\ 2 & 10 & 1 \\ 3 & 1 & 2 \end{pmatrix} \text{ and } b = \begin{pmatrix} 7 \\ -16 \\ 5 \end{pmatrix}.$$

$$A = QR = \begin{pmatrix} -0.81 & 0.27 & -0.52 \\ -0.32 & -0.95 & 0.02 \\ -0.49 & 0.18 & 0.85 \end{pmatrix} \begin{pmatrix} -6.16 & -5.35 & -3.73 \\ 0 & -8.74 & 0.23 \\ 0 & 0 & 0.17 \end{pmatrix}$$

$$Q^T b = \begin{pmatrix} -2.92 \\ 17.93 \\ 0.33 \end{pmatrix} \quad \text{and} \quad Rx = Q^T b \quad \text{gives} \quad x = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}.$$

Computing the QRD

- Givens Rotations
- Householder transformations
- Gram-Schmidt process

Singular Value Decomposition (SVD)

Let $A \in \Re^{m \times n}$ be a matrix of rank k.
The Singular Value Decomposition (SVD) of A is given by:

$$\boxed{A = Q \Sigma P^T},$$

- where $Q \in \Re^{m \times m}$ and $P \in \Re^{n \times n}$ are orthogonal,
- $\Sigma = \begin{pmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$

SVD

- $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0 \quad$ and $\quad \sigma_{k+1} = \ldots \sigma_n = 0.$
- The rank of A is k.
- The σ_i is called the ith singular value of A.
- If $Q = (q_1, \ldots q_m)$ and $P = (p_1, \ldots p_n)$, then q_i and p_i are called the left and right singular vectors associated with σ_i $(i = 1, \ldots, k)$.
- The ratio $\kappa(A) = \sigma_1/\sigma_n$ is called the condition number of A.

SVD. Example

$$A = \begin{pmatrix} -6 & -12 & 8 \\ 2 & 12 & -11 \\ -6 & -17 & 10 \\ 19 & 3 & 6 \\ -9 & 6 & 15 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} 31.71 & 0 & 0 \\ 0 & 19.80 & 0 \\ 0 & 0 & 16.99 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$Q = \begin{pmatrix} -0.49 & -0.05 & -0.11 & 0.51 & -0.70 \\ 0.49 & 0.29 & 0.02 & 0.80 & 0.22 \\ -0.63 & -0.15 & -0.23 & 0.26 & 0.68 \\ 0.23 & -0.93 & 0.20 & 0.19 & 0.01 \\ -0.27 & 0.15 & 0.94 & 0.06 & 0.09 \end{pmatrix}, P = \begin{pmatrix} 0.46 & -0.88 & -0.15 \\ 0.68 & 0.23 & 0.70 \\ -0.58 & -0.42 & 0.70 \end{pmatrix}.$$

The Condition number of A is given by $\sigma_1/\sigma_3 = 31.77/16.99 = 1.87$.
Consider the matrices:

$$A_0 = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 1 \\ 3 & 6 & 0 \\ 4 & 8 & 2 \end{pmatrix}, A_1 = \begin{pmatrix} 1 & 2.01 & 0 \\ 2 & 3.99 & 1 \\ 3 & 6 & 0 \\ 4 & 8 & 2 \end{pmatrix}, A_2 = \begin{pmatrix} 1 & 2.1 & 0 \\ 2 & 3.9 & 1 \\ 3 & 6 & 0 \\ 4 & 8 & 2 \end{pmatrix}, A_3 = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 4 & 1 \\ 3 & 9 & 0 \\ 4 & 16 & 2 \end{pmatrix}.$$

Cond(A_0)=8.82e+16, Cond(A_1)=2124.5, Cond(A_2)=213.02,
Cond(A_3)=17.77, and Cond(I_n)=1.

Content

The Eigenvalue problem

- Let A be a square matrix of order $n \times n$, $x \neq 0$ is an n-element column vector and λ is a scalar.
- The EIGENVALUE PROBLEM: Solve

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

- The solution come in pairs: to each λ corresponds an x vector.
- The λ's are known as eigenvalues (or latent, or characteristic roots).
- The x's as eigenvectors (or latent, or, characteristic vectors).

The Eigenvalue problem

- In matrix format the Eigenvalue problem can be written as:

$$\boxed{(A - \lambda I_n)x = 0}$$

- In order for $x \neq 0$ it implies that

$$\boxed{|A - \lambda I_n| = 0}.$$

- The latter is known as the *characteristic equation* for A. It gives a polynomial equation in the unknown λ.

Example

Let $A = \begin{pmatrix} 1 & 0 \\ 1 & 3 \end{pmatrix}$ so that $A - \lambda I_2 = \begin{pmatrix} 1-\lambda & 0 \\ 1 & 3-\lambda \end{pmatrix}$.
Now, $|A - \lambda I_2| = (1-\lambda)(3-\lambda)$.
Thus, $\lambda_1 = 1$ and $\lambda_2 = 3$ are the eigenvalues of A.

For the eigenvalue $\lambda_1 = 1$ we have $Ax = \lambda_1 x$:

$$\begin{pmatrix} 1 & 0 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \text{or} \quad \begin{aligned} x_1 &= x_1 \\ x_1 &= -2x_2. \end{aligned}$$

Thus, an eigenvector of A corresponding to the eigenvalue $\lambda_1 = 1$ is given by $x = (-2 \quad 1)^T$. Normalizing x, i.e. dividing each of its entries by $\sqrt{x^T x}$, it gives the eigenvector

$$\frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

An eigenvector associated with the eigenvalue of $\lambda_2 = 3$ is given by $(0 \quad 1)^T$.

Properties of eigenvalues and eigenvectors

Given an $m \times m$ SYMMETRIC matrix, e.g. the variance-covariance matrix: $A = \begin{pmatrix} 5 & 2 & 3 \\ 2 & 10 & 1 \\ 3 & 1 & 2 \end{pmatrix}$

- **The eigenvalues are real.** The eigenvalue of A are given by:
  $\lambda_1 = 0.14, \quad \lambda_2 = 5.70 \quad \text{and} \quad \lambda_3 = 11.16.$

Properties of eigenvalues and eigenvectors

- **Eigenvectors corresponding to dinstinct eigenvalues are pairwise orthogonal**[1]. I.e. if x_1 and x_2 are the eigenvectors corresponding to the eigenvalues λ_1 and λ_2 $(\lambda_1 \neq \lambda_2)$, then $x_1^T x_2 = 0$.

  The eigenvectors of A are given by the columns of $X = (x_1, x_2, x_3)$, where

$$X = \begin{pmatrix} -0.532 & 0.747 & 0.400 \\ 0.022 & -0.459 & 0.888 \\ 0.847 & 0.481 & 0.228 \end{pmatrix} \quad \text{and} \quad X^T X = XX^T = I_3.$$

[1] Notice that $Ax_1 = \lambda_1 x_1$, and after premultiplicationn by x_2^T it gives $x_2^T A x_1 = \lambda_1 x_2^T x_1$. Similarly, $x_1^T A x_2 = \lambda_2 x_1^T x_2$. Since $x_2^T A x_1 = x_1^T A x_2$ it follows that $\lambda_1 x_2^T x_1 = \lambda_2 x_1^T x_2$ and thus, $x_1^T x_2 = 0$.

Properties of eigenvalues and eigenvectors

- **The orthogonal matrix of eigenvectors diagonalizes**[2] A. That is,

$$\boxed{X^T A X = \Lambda},$$

where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_m)$ and $X = (x_1 \dots x_m)$.

$$X^T A X = \begin{pmatrix} 0.14 & 0 & 0 \\ 0 & 5.70 & 0 \\ 0 & 0 & 11.16 \end{pmatrix} = \Lambda$$

- **The matrices A and A^T have the same eigenvalues.**
- **The matrix A is singular if one of its eigenvalues is zero.**
- **The rank of A is equal to the number of non-zero eigenvalues.**

[2] The Eigenvalue problem in matrix form is equivalent to $AX = X\Lambda$. Premultiplying by X^T it gives $X^T A X = X^T X \Lambda$ which is equivalent to $X^T A X = \Lambda$ since $X^T X = I_m$.

Properties of eigenvalues and eigenvectors

- $A^2 = AA = X\Lambda^2 X^T$ and generally $A^n = X\Lambda^n X^T$.

$$A^2 = \begin{pmatrix} 38 & 33 & 23 \\ 33 & 105 & 18 \\ 23 & 18 & 14 \end{pmatrix} \quad \text{and} \quad \Lambda^2 = \begin{pmatrix} 0.02 & 0 & 0 \\ 0 & 32.51 & 0 \\ 0 & 0 & 124.47 \end{pmatrix}.$$

- $A^{-1} = X\Lambda^{-1} X^T$ since $(X\Lambda X^T)^{-1} = X\Lambda^{-1} X^T$.

$$A^{-1} = \frac{1}{9}\begin{pmatrix} 19 & -1 & -28 \\ -1 & 1 & 1 \\ -28 & 1 & 46 \end{pmatrix} \text{ and } \Lambda^{-1} = \begin{pmatrix} 7.07 & 0 & 0 \\ 0 & 0.18 & 0 \\ 0 & 0 & 0.09 \end{pmatrix}.$$

Equiv. of the SVD of A and Eigensystem of $A^T A$

Consider the SVD of $A \in \Re^{m \times n}$: $\quad A = Q\Sigma P^T$, where Q and P have orthogonal columns, and $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n)$.
Now, $\quad A^T A = (P\Sigma Q^T)(Q\Sigma P^T) = P\Sigma^2 P^T$, or

$$\boxed{P^T A^T A P = \Sigma^2}.$$

Thus, the SVD of A provides:

- The eigenvectors P of the symmetric $A^T A$
- The diagonal elements of Σ are the positive square roots of the eigenvalues of $A^T A$. I.e. $\lambda_1 = \sigma_1^2, \ldots, \lambda_n = \sigma_n^2$.

Equiv. of the SVD of A and Eigensystem of $A^T A$

Let $A = \begin{pmatrix} 5 & 2 & 3 \\ 2 & 10 & 1 \\ 3 & 1 & 2 \end{pmatrix}$ such that $A^T A = \begin{pmatrix} 38 & 33 & 23 \\ 33 & 105 & 18 \\ 23 & 18 & 14 \end{pmatrix}$.

The singular values of A are: $\sigma_1 = 11.16$, $\sigma_2 = 5.70$ and $\sigma_3 = 0.14$.

The eigenvalues of $A^T A$ are: $\lambda_1 = 124.47$, $\lambda_2 = 32.51$ and $\lambda_3 = 0.02$.

Quadratic forms and definite matrices

Consider the quadratic form $q = x^T A x$, where A is a symmetric matrix and $x \neq 0$. E.g. if $A \in \Re^{2\times 2}$, then

$$\boxed{q = x^T A x = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2}.$$

- If $x^T A x > 0$, then the quadratic form is said to be positive definite. *In this case all the eigenvalues of A are positive.*
  E.g. Let $S = \begin{pmatrix} 5 & 2 \\ 2 & 10 \end{pmatrix}$ such that $\Lambda = \begin{pmatrix} 10.70 & 0 \\ 0 & 4.29 \end{pmatrix}$.

Quadratic forms and definite matrices

- If $x^T Ax \geq 0$, then the quadratic form is said to be positive (or nonnegative) semidefinite. *In this case all the eigenvalues of A are positive or zero.*
  E.g. Let $S = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ such that $\Lambda = \begin{pmatrix} 5 & 0 \\ 0 & 0 \end{pmatrix}$.

- If $x^T Ax < 0$, then the quadratic form is said to be negative definite. *In this case all the eigenvalues of A are negative.*

- If $x^T Ax \leq 0$, then the quadratic form is said to be negative (or nonpositive) semidefinite. *In this case all the eigenvalues of A are negative or zero.*

Content

Kronecker products

- A calculation that helps condense the notation when dealing with sets of regression models are the *Kronecker product* and *vector operator*.
- The KRONECKER PRODUCT of the two matrices $A = [a_{ij}] \in \Re^{m\times n}$ and $B = [b_{ij}] \in \Re^{p\times q}$ is defined by:

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{pmatrix}$$

- Notice that $A \otimes B$ has dimension $mp \times nq$.

Kronecker products. Example

Let $A = \begin{pmatrix} 3 & 0 \\ 5 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 4 \\ -1 & 0 \\ -2 & 1 \end{pmatrix}$:

$$A \otimes B = \begin{pmatrix} 3\begin{pmatrix} 1 & 4 \\ -1 & 0 \\ -2 & 1 \end{pmatrix} & 0\begin{pmatrix} 1 & 4 \\ -1 & 0 \\ -2 & 1 \end{pmatrix} \\ 5\begin{pmatrix} 1 & 4 \\ -1 & 0 \\ -2 & 1 \end{pmatrix} & 4\begin{pmatrix} 1 & 4 \\ -1 & 0 \\ -2 & 1 \end{pmatrix} \end{pmatrix} = \left(\begin{array}{rr|rr} 3 & 12 & 0 & 0 \\ -3 & 0 & -0 & 0 \\ -6 & 3 & -0 & 0 \\ \hline 5 & 20 & 2 & 8 \\ -5 & 0 & -2 & 0 \\ -10 & 5 & -4 & 2 \end{array}\right).$$

$$A \otimes I_2 = \begin{pmatrix} 3\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & 0\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ 5\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & 2\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{pmatrix} = \left(\begin{array}{cc|cc} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ \hline 5 & 0 & 2 & 0 \\ 0 & 5 & 0 & 2 \end{array}\right).$$

$$I_2 \otimes A = \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix} = \left(\begin{array}{cc|cc} 3 & 0 & 0 & 0 \\ 5 & 2 & 0 & 0 \\ \hline 0 & 0 & 3 & 0 \\ 0 & 0 & 5 & 2 \end{array}\right).$$

Direct sum of matrices

- Given the set of matrices $\{A_1, \ldots, A_G\}$ the DIRECT SUM of matrices is defined by:

$$\bigoplus_{i=1}^{G} A_i = \mathrm{diag}(A_1, \ldots, A_G) = \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_G \end{pmatrix}.$$

- Notice that the matrices $A_1, \ldots, A_G$ can have different dimensions.

- In the event where the matrices are of the same (A) then:

$$\bigoplus_{i=1}^{G} A_i = I_G \otimes A.$$

Vector operator

Let the $m \times n$ matrix $Y = (y_1 \dots y_n)$ where $y_i \in \Re^m$ is the ith column of Y. The vec$(\cdot)$ operator stacks the columns of Y one under the other. That is,

$$\boxed{\mathrm{vec}(Y) = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}.$$

E.g. If $Y = \begin{pmatrix} 1 & 4 \\ 1 & 0 \\ 2 & 1 \end{pmatrix}$, then $\mathrm{vec}(Y) = \begin{pmatrix} 1 & 1 & 2 & 4 & 0 & 1 \end{pmatrix}^T$.

Vector operator

Given the set of vectors $\{y_i\}_G = \{y_1, \ldots, y_G\}$ the $\text{vec}(\cdot)$ operator stacks the vectors one under the other:

$$\boxed{\{\text{vec}(y_i)_G\} = \begin{pmatrix} y_1 \\ \vdots \\ y_G \end{pmatrix}}.$$

Properties

Assuming appropriate dimensions the following properties exist:

- $(A \otimes B)(C \otimes D) = (AC \otimes BD)$.
- $(A \otimes B)^T = (A^T \otimes B^T)$.
- $(A \otimes B)^{-1} = (A^{-1} \otimes B^{-1})$.
- $|A \otimes B| = |A|^M |B|^N$.
- $(A \otimes B)\mathrm{vec}(C) = \mathrm{vec}(BCA^T)$.

Vector and Frobenius norms

The p-norm of $x \in \Re^n$ is defined as:

$$\boxed{\|x\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{\frac{1}{p}}, \quad \text{where} \quad p \geq 1.}$$

Important norms:

- $\|x\|_1 = (|x_1| + \cdots + |x_n|)$.
- $\|x\|_2 = (|x_1|^2 + \cdots + |x_n|^2)^{\frac{1}{2}} = \sqrt{x^T x}$.
  The 2-norm is also known as EUCLIDIAN NORM.
  Hereafter $\|\cdot\|$ will denote the Euclidian norm.
- $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$.

Given a matrix $A = [a_{ij}] \in \Re^{m \times n}$ the FROBENIUS NORM of A is given by:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}.$$

Absolute and relative error

Suppose $\hat{x} \in \Re^n$ is an approximation of $x \in \Re^n$. Then:

- Absolute error in $\hat{x}$: $\|\hat{x} - x\|$.
- Relative error in $\hat{x}$: $\|\hat{x} - x\| \Big/ \|x\|$.

Content

Random vectors and matrices

A RANDOM VECTOR (RANDOM MATRIX) is a vector (matrix) whose elements are random variables. The expected value of a random matrix consists of the expected values of each element. That is, if $X = [X_{ij}] \in \Re^{m \times n}$, then

$$\boxed{E(X) = \begin{pmatrix} E(X_{11}) & \dots & E(X_{1n}) \\ \vdots & & \vdots \\ E(X_{m1}) & \dots & E(X_{mn}) \end{pmatrix}},$$

where

$$E(X_{ij}) = \begin{cases} \int_{-\infty}^{\infty} x_{ij} f_{ij}(x_{ij}) dx_{ij} & \text{If } X_{ij} \text{ is continues random variable with pdf } f_{ij}(x_{ij}) \\ \sum_{\text{all } x_{ij}} x_{ij} p_{ij}(x_{ij}) & \text{If } X_{ij} \text{ is discrete r.v. with probability function } p_{ij}(x_{ij}) \end{cases}$$

Example

Consider the random vector $X^T = (X_1 \quad X_2)$, where X_1 and X_2 have, respectively, the following probability functions:

x_1	-1	0	1
$p_1(x_1)$	0.3	0.3	0.4

x_2	0	1
$p_2(x_2)$	0.8	0.2

Thus, $E(X) = (E(X_1) \quad E(X_2))^T = (0.1 \quad 0.2)^T$.

Mean vectors and covariance matrices

Consider the random vector $X = [X_i] \in \Re^n$, where X_i has mean $\mu_i = E(X_i)$ and variance $\sigma_i^2 = E(X_i - \mu_i)^2$, where $i = 1, \ldots, n$.

The means of the vector X is given by:

$$\boxed{E(X) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = \mu.}$$

Mean vectors and covariance matrices

Specifically,

$$\mu_i = \begin{cases} \int_{-\infty}^{\infty} x_i f_i(x_i) dx_i & \text{if } X_i \text{ is continues random variable} \\ & \text{with pdf } f_i(x_i) \\ \sum_{\text{all } x_i} x_i p_i(x_i) & \text{if } X_i \text{ is discrete random variable} \\ & \text{with probility function } p_i(x_i) \end{cases}$$

and

$$\sigma_i^2 = \begin{cases} \int_{-\infty}^{\infty} (x_i - \mu_i)^2 f_i(x_i) dx_i & \text{if } X_i \text{ is continues random variable} \\ & \text{with pdf } f_i(x_i) \\ \sum_{\text{all } x_i} (x_i - \mu_i)^2 p_i(x_i) & \text{if } X_i \text{ is discrete random variable} \\ & \text{with probility function } p_i(x_i) \end{cases}$$

Mean vectors and covariance matrices

- The behavior of any pair of random variables, such as X_i and X_k, is described by their joint probability function and a the covariance σ_{ik}.

- The covariance σ_{ik} is measure of a linear associassion between the two variables and is given by:

$$\boxed{\sigma_{ik} = E(X_i - \mu_i)(X_k - \mu_k).}$$

Mean vectors and covariance matrices

If X_i and X_k are continues random variables with joint density function $f_{ik}(x_i, x_k)$, then

$$\boxed{\sigma_{ik} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_i - \mu_i)(x_k - \mu_k) f_{ik}(x_i, x_k)dx_i dx_k.}$$

If X_i and X_k are discrete random variables with joint probability function $p_{ik}(x_i, x_k)$, then

$$\boxed{\sigma_{ik} = \sum_{\text{all } x_i}\sum_{\text{all } x_k}(x_i - \mu_i)(x_k - \mu_k) p_{ik}(x_i, x_k).}$$

Generally, the collective behavior of the n random variables $X_1, \ldots, X_n$, or equivalently the random vector $X^T = (X_1 \ldots X_n)$ is described by the a joint probability density function $f(x_1, \ldots, x_n) = f(x)$ and their covariance matrix.

Mean vectors and covariance matrices

The covariances of the vector X is given by:

$$Var(X) = \Sigma = E(X-\mu)(X-\mu)^T$$

$$= \mathrm{E}\begin{pmatrix} (X_1-\mu_1)^2 & (X_1-\mu_1)(X_2-\mu_2) & \dots & (X_1-\mu_1)(X_n-\mu_n) \\ (X_2-\mu_2)(X_1-\mu_1) & (X_2-\mu_2)^2 & \dots & (X_2-\mu_2)(X_n-\mu_n) \\ \vdots & \vdots & \ddots & \vdots \\ (X_n-\mu_n)(X_1-\mu_1) & (X_n-\mu_n)(X_2-\mu_2) & \dots & (X_n-\mu_n)^2 \end{pmatrix}$$

$$= \begin{pmatrix} \mathrm{E}(X_1-\mu_1)^2 & \mathrm{E}(X_1-\mu_1)(X_2-\mu_2) & \dots & \mathrm{E}(X_1-\mu_1)(X_n-\mu_n) \\ \mathrm{E}(X_2-\mu_2)(X_1-\mu_1) & \mathrm{E}(X_2-\mu_2)^2 & \dots & \mathrm{E}(X_2-\mu_2)(X_n-\mu_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}(X_n-\mu_n)(X_1-\mu_1) & \mathrm{E}(X_n-\mu_n)(X_2-\mu_2) & \dots & \mathrm{E}(X_n-\mu_n)^2 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{pmatrix}.$$

Example

Find the covariance matrix for the two random variables X_1 and X_2 when their joint probability function $p_{12}(x_1, x_2)$ is represented by following table:

x_1 \ x_2	0	1	$p_1(x_1)$
-1	0.24	0.06	0.3
0	0.16	0.14	0.3
1	0.40	0.00	0.4
$p_2(x_1)$	0.8	0.2	1

Notice that: $\mu_1 = E(X_1) = 0.1$ and $\mu_2 = E(X_2) = 0.2$.

Example

$$\sigma_{11} = E(X_1 - \mu_1)^2 = \sum_{\text{all } x_1} (x_i - 0.1)^2 p_1(x_1)$$

$$= (-1 - .1)^2(.3) + (0 - .1)^2(.3) + (1 - .1)^2(.4) = 0.69$$

$$\sigma_{22} = E(X_2 - \mu_2)^2 = \sum_{\text{all } x_2} (x_i - 0.2)^2 p_2(x_2)$$

$$= (0 - .2)^2(.8) + (1 - .2)^2(.2) = .16$$

$$\sigma_{12} = E(X_1 - \mu_1)(X_2 - \mu_2)$$

$$= \sum_{\text{all pairs } (x_1, x_2)} (x_1 - 0.1)(x_2 - 0.2) p_{12}(x_1, x_2)$$

$$= (-1 - .1)(0 - .2)(.24) + (-1 - .1)(1 - .2)(.06) + \cdots = -0.08$$

$$\sigma_{21} = E(X_2 - \mu_2)(X_1 - \mu_1) = \sigma_{12} = -0.08.$$

Thus, the mean and covariance matrix of X are given, respectively, by:

$$\mu = E(X) = \begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix} \text{ and } Var(X) = \Sigma = \begin{pmatrix} 0.69 & -0.08 \\ -0.08 & 0.16 \end{pmatrix}.$$

Multivariate Normal

The n random variables $X^T = (X_1, \ldots, X_n)$ have some probability density function (pdf) which is written as:

$$p(X) = p(X_1, \ldots, X_n).$$

This gives the likelihood of various combinations of X values. The most important multivariate pdf is the multivariate normal. It is specified in terms of its mean vector μ and its variance matrix Σ. The formula of the multivariate normal is given by:

$$\boxed{p(X) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(X-\mu)^T \Sigma^{-1}(X-\mu)\right).}$$

Compactly the latter is stated as:

$$\boxed{X \sim N(\mu, \Sigma).}$$

Properties

Let X and Y be random matrices of the same dimension and let A and B be conformable matrices of constants. Then,

- $E(X+Y) = E(X) + E(Y)$.
- $E(AXB) = A\,E(X)\,B$.
- $Var(AX) = A\,Var(X)\,A^T$.

Example

Let $X \in \Re^n$ have a positive definite covariance matrix Σ. Furthermore, let the Cholesky factor of $\Sigma = CC^T$. Find the covariance matrix of $Z = C^{-1}X$.

$$\begin{aligned} Var(Z) &= Var(C^{-1}X) = C^{-1}\,Var(X)\,C^{-T} \\ &= C^{-1}\Sigma C^{-T} = C^{-1}CC^TC^{-T} \\ &= I_n. \end{aligned}$$