Fundamentos para al Análisis de Datos y la Investigación

Tema 3 - Fundamentos de Matemáticas para la Ciencia de Datos (1/3)

José María Sarabia

Máster Universitario en Ciencia de Datos
CUNEF Universidad

Contents

1. Introducción
2. Operaciones con matrices
3. Aplicaciones en finanzas
4. Dependencia y rango en matrices
5. Aplicaciones en CD

1 Introducción

Fundamentos de álgebra en Ciencia de Datos. Contenidos

- Matrices. Tipos y operaciones con matrices. Matrices especiales.
- Aplicaciones en finanzas: media y varianza de una cartera de valores
- Rango de una matriz. Dependencia lineal. Matriz inversa. Determinantes
- Casos Prácticos en R

1 Introduction

Objetivos

- Revisar conceptos de álgebra en ciencia de datos
- Revisar y manejar los diferentes tipos y operaciones con matrices.
- Revisar el concepto de dependencia lineal y rango de una matriz. Obtener la matriz inversa y el determinante de una matriz.
- Saber realizar operaciones de matrices con el software R

Contents

1 Introducción

2 Operaciones con matrices

3 Aplicaciones en finanzas

4 Dependencia y rango en matrices

5 Aplicaciones en CD

1 Vectores

- Los vectores son filas o columnas de datos. Normalmente se representan por letras minúsculas.
- En análisis de datos, los vectores son columnas de n datos es decir, con dimensión $n \times 1$
- Un vector 3×1 es:

$$\mathbf{v} = \begin{pmatrix} 1 \\ 5 \\ 2 \end{pmatrix}$$

- Los vectores tienen una representación gráfica y un vector $n \times 1$ es un punto en un espacio de dimensión n

1 Matrices

- Una matriz es una disposición de números en forma de rectángulo. Las matrices se denotarán letras mayúsculas en negrita $\mathbf{A}$, $\mathbf{B}$ etc.
- La dimensión de una matriz (orden) es el número de filas (rows) y columnas (columns), y se escribe $r \times c$. Los elementos de una matriz se identifican por su posición en la fila y la columna.
- La transpuesta de una matriz $\mathbf{A}$ $r \times c$ es el resultado de cambiar en $\mathbf{A}$ las filas por las columnas, y se denota por $\mathbf{A}^\top$ y su dimensión es $c \times r$. Si $\mathbf{A} = \{a_{ij}\}$ entonces $\mathbf{A}^\top = \{a_{ji}\}$
- La transpuesta de un vector columna es un vector fila y viceversa.

1 Matrices en R

- La matrix 3×3

$$\mathbf{A} = \begin{pmatrix} 3 & 5 & 8 \\ 5 & 7 & 1 \\ -4 & 8 & -3 \end{pmatrix}$$

Se representa en `R` como:

```
A <- matrix(c(3,5,-4,5,7,8,8,1,-3), ncol=3,nrow=3)
```

1 Operaciones con matrices

- Las matrices pueden sumarse y restarse únicamente si son del mismo orden.
$$\mathbf{A}+\mathbf{B}=(a_{ij}+b_{ij}), \text{ y } \mathbf{A}-\mathbf{B}=(a_{ij}-b_{ij})$$
- Producto de una matriz **A** por un escalar λ
$$\lambda\mathbf{A}=(\lambda a_{ij})$$
- Las matrices pueden mutiplicarse sólo si los ordener son compatibles Si **A** es de orden $r \times c$ y **B** es de orden $c \times d$ entonces **AB** tiene orden $r \times d$.
- Las matrices **no** se pueden dividir. Si se da la compatibilidad de órdenes, entonces se puede obtener el producto de **A** por $\mathbf{B}^{-1}$. Si los órdenes son compatibles, entonces se puede obtener $\mathbf{A}\mathbf{B}^{-1}$ o bien $\mathbf{B}^{-1}\mathbf{A}$, si existe la inversa.

1 Producto de matrices

- El producto de dos matrices es otra matriz donde cada elemento es el producto de fila por columna
- Si $\mathbf{A} = \{a_{ij}\}$ es $m \times n$ y $\mathbf{B} = \{b_{ij}\}$ es $n \times p$, entonces $\mathbf{C} = \mathbf{AB}$ es $m \times p$, donde $\mathbf{C} = \{c_{ij}\} = \{\sum_{r=1}^{n} a_{ir} b_{rj}\}$
- En general $\mathbf{AB} \neq \mathbf{BA}$, salvo para algunas matrices especiales.

1 Matrices especiales

Las siguientes matrices son todas cuadradas:

- *Matrices diagonales*: únicamente los elementos de la diagonal pueden ser diferentes de cero
- *Matrices simétricas*: $\mathbf{A} = \mathbf{A}^\top$
- *Matrices invertibles* o matrices no singulares, son aquellas que $\mathbf{A}^{-1}$ existe
- *Matrices ortogonales*: son aquellas que verifican: $\mathbf{A}^{-1} = \mathbf{A}^\top$
- *Matrices triangulares* superior (inferior): on aquellas donde únicamante los elementos por encima de la diagonal (debajo) son distintos de cero.

1 Matriz identidad

- La matriz identidad, $\mathbf{I}$ es una matriz diagonal que contiene sólo unos.
- La matriz identidad 2×2 es:

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

- La matriz identidad 3×3 es:

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- Para cualquier matriz cuadrada $\mathbf{A}$:

$$\mathbf{AI} = \mathbf{IA} = \mathbf{A}$$

Contents

1. Introducción
2. Operaciones con matrices
3. **Aplicaciones en finanzas**
4. Dependencia y rango en matrices
5. Aplicaciones en CD

2 Aplicaciones en finanzas

Matrices de covarianzas y de correlaciones

- Disponemos de un conjunta de variables $X_1, \ldots, X_n$ (por ejemplo rendimientos de un conjunto de n activos)
- La matriz de covarianzas $\mathbf{V}$ es una matriz $n \times n$ simétrica, donde $v_{ij} = \text{cov}(X_i, X_j)$ si $i \neq j$ y en la diagonal $v_{ii} = \text{cov}(X_i, X_i) = \text{var}(X_i)$
- La matriz de correlaciones $\mathbf{R}$ es una matriz simétrica $n \times n$ donde $r_{ij} = \text{cor}(X_i, X_j)$ si $i \neq j$ y $r_{ii} = \text{cor}(X_i, X_i) = 1$
- La matriz de covarianzas $\mathbf{V}$ se puede escribir como el producto:

$$\mathbf{V} = \mathbf{DRD}$$

donde $\mathbf{D}$ es la matriz diagonal de las desviaciones típicas.

2 Rendimiento de una cartera

Media y varianza (riesgo) del rendimiento de una cartera

- Tenemos una cartera con k activos con riesgo con rendimientos $\mathbf{r} = (R_1, \ldots, R_k)$ y pesos $\mathbf{w} = (w_1, \ldots, w_k)$. El rendimiento de la cartera se puede expresar como

$$R = \sum_{i=1}^{k} w_i R_i \Leftrightarrow R = \mathbf{w}^\top \mathbf{r}$$

- El rendimiento medio de la cartera es

$$E(R) = w_1 E(R_1) + \cdots + w_k E(R_k) = \sum_{i=1}^{k} w_i E(R_i) = \mathbf{w}^\top E(\mathbf{r})$$

- La matriz de covarianza de los activos $\mathbf{V}$ es una matriz $k \times k$ simétrica semidefinida positiva donde el elementos (i,j) es igual a $\mathrm{cov}(R_i, R_j)$

2 Riesgo y rendimiento de una cartera

- La varianza del rendimiento de la cartera es:

$$\mathrm{var}(R) = \sum_{i=1}^{k} w_i^2 var(R_i) + \sum_{i=1}^{k}\sum_{j=1}^{k} w_i w_j \mathrm{cov}(R_i, R_j)$$

  y en notación matricial:

$$\mathrm{var}(R) = \mathbf{w}^\top \mathbf{V}\mathbf{w}$$

- Si $\mathbf{x} = \mathbf{D}\mathbf{w}$ donde $\mathbf{D}$ es la matriz diagonal de las volatilidades (desviaciones típicas) de los activos:

$$\mathrm{var}(R) = \mathbf{x}^\top \mathbf{C}\mathbf{x}$$

  donde $\mathbf{C}$ es la matriz de correlaciones.

Contents

1 Introducción

2 Operaciones con matrices

3 Aplicaciones en finanzas

4 Dependencia y rango en matrices

5 Aplicaciones en CD

3 Dependencia lineal

- *La dependencia lineal* se refiere al caso donde existe una relación lineal entre filas o columnas de una matriz
- Cuando existe dependencia lineal, se dice que tenemos una matriz *singular*, porque no tiene inversa (matrices cuadradas).
- En general, el número de filas (o columnas) linealmente independientes se llama *rango* de la matriz.
- Si todas las filas (o columnas) son linealmente independientes, se dice que la matriz tiene *rango completo*.
- Las matrices de rango completo son *no singulares*

3 Rango de una matriz

- *El rango* de una matriz es el número máximo de filas o columnas linealmente independientes.
- La forma más simple de obtener el rango de una matriz `M` es haciendo `qr(M)$rank`
- Otra opción es cargar la librería `Matrix` y usar la función `rankMatrix(M)`

3 Matriz inversa

- La *inversa de una matriz* $\mathbf{A}$ es la matriz $\mathbf{A}^{-1}$ tal que

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

- Por tanto, el producto de una matriz por su inversa es la matriz identidad
- No todas las matrices tienen inversa, pero si existe es única
- Unicamente las matrices no singulares tienen inversa
- Un test rápido para sober que una matriz es no singular (es decir, es invertible) es probar que su determinante no es cero.

3 Determinantes

- Para una matriz cuadrada 2×2 la cantidad $ad - bc$ se denomina *determinante* de la matriz.
- Se suele representar:

$$\det(\mathbf{A}) = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

- El determinante de una matriz 3×3 es

$$\begin{vmatrix} a & b & c \\ u & v & w \\ x & y & z \end{vmatrix} = a \begin{vmatrix} v & w \\ y & z \end{vmatrix} - b \begin{vmatrix} u & w \\ x & z \end{vmatrix} + c \begin{vmatrix} u & v \\ x & y \end{vmatrix}$$

3 Propiedades de los determinantes

- Los determinantes sólo existen para matrices cuadradas
- Las siguientes propiedades son útiles para el cálculo de determinantes:
  - Para cualquier matriz cuadrada: $|\mathbf{A}^\top| = |\mathbf{A}|$
  - Para cualquier matriz invertible: $|\mathbf{A}^{-1}| = |\mathbf{A}|^{-1}$
  - Para una constante c y una matriz $\mathbf{A}$ $n \times n$: $|c\mathbf{A}| = c^n|\mathbf{A}|$
  - Si A y B son matrices $n \times n$: $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$

3 Operaciones con matrices en R

Operador	Descripción
`A+B`	suma de matrices
`A*B`	producto elemento a elemento
`A%*%B`	producto de matrices
`crossprod(A,B)`	$A^\top B$
`crossprod(A)`	$A^\top A$
`t(A)`	traspuesta de A
`diag(x)`	matriz diagonal con elementos x en la diagonal principal
`diag(A)`	con elementos la diagonal de A
`diag(k)`	matriz identidad $k \times k$, con k entero
`solve(A, b)`	vector x solución de $Ax = b$ (i.e., $x = A^{-1}b$)
`solve(A)`	matriz inversa de A donde A es matriz cuadrada
`ginv(A)`	matriz inversa generalizada de Moore-Penrose de A, con MASS
`cbind(A,B,...)`	combina matrices o vectores horizon. y devuelve una matriz
`rbind(A,B,...)`	combina matrices o vectores vert. y devuelve una matriz
`rowMeans(A)`	vector de media de filas
`rowSums(A)`	vector suma de filas
`colMeans(A)`	vector medias de columnas
`colSums(A)`	vector suma de columnas

Contents

1 Introducción

2 Operaciones con matrices

3 Aplicaciones en finanzas

4 Dependencia y rango en matrices

5 Aplicaciones en CD

4 Aplicaciones en Ciencia de Datos

- *Matrices*: de covarianzas y de correlaciones, basadas en distancias, etc.
- *Manipulación y transformaciones de datos*
- *Análisis de regresión*
- *Reducción de la dimension*:
  - Análisis de componentes principales (PCA)
  - Análisis de componentes principales robusto
  - Análisis factorial

Fundamentos para al Análisis de Datos y la Investigación

Tema 3 - Fundamentos de Matemáticas para la Ciencia de Datos (1/3)

José María Sarabia

Máster Universitario en Ciencia de Datos
CUNEF Universidad