EST-46114 Métodos Multivariados y Datos Categóricos

Sesion 01 - Algebra Lineal

Juan Carlos Martinez-Ovando

Maestria en Ciencia de Datos

Objetivo

Revisar y entender los aspectos teóricos y prácticos asociados con los operadores algebraicos empleados en modelos estadísticos unidimensionales, en particular formas cuadráticas, inversión, trazas, rangos y descomposición singular.

1. Propiedades básicas

Una *matriz* es un arreglo numérico bidimensional definido con entradas numéricas, dado por

$$\boldsymbol{A} = (a_{ij})_{i=1,j=1}^{m,p} = (a_{i\cdot})_{i=1}^{m} = (b_{\cdot j})_{j=1}^{p},$$

donde $a_{i\cdot}$ es un vector de dimensión $p \times 1$ y $b_{\cdot j}$ es un vector de dimensión $m \times 1$.

- Se dice que la dimensión de $\boldsymbol{A}$ es $m \times p$.

La *matriz transpuesta* de $\boldsymbol{A}$ se define como,

$$\boldsymbol{A}' := (a_{j,i})_{j=1,i=1}^{p,m}.$$

1.0. Productos

El *producto* de dos matrices $\boldsymbol{A}$ y $\boldsymbol{B}$ de dimensiones $m \times p$ y $p \times n$, respectivamente, se define como,

$$\boldsymbol{AB} = (a_{i.}' b_{.j})_{i=1,j=1}^{m,n},$$

donde $a_{i.}' b_{.j}$ denota el producto interior entre los vectores $a_{i.}$, de dimensión $p \times 1$, y $b_{.j}$, de dimensión $p \times 1$.

1.1. Trazas

La traza de una matriz se define como la suma de los elementos de su diagonal, i.e.

$$tr(\boldsymbol{A}) = \begin{cases} \sum_{i=1}^{m} a_{ii} & \text{para } m = p, \\ \text{no definda} & \text{para } m \neq p. \end{cases}$$

Es decir, la traza es un operador válido para matrices cuadradas (mismo número de renglones y columnas) solamente.
El operador $tr(\cdot)$ tiene varias propiedades interesantes, entre ellas:

- $tr(\boldsymbol{A}) = tr(\boldsymbol{A}')$.
- $tr(\boldsymbol{AB}) = tr(\boldsymbol{BA})$.
- $tr(\boldsymbol{A} + \boldsymbol{B}) = tr(\boldsymbol{B}) + tr(\boldsymbol{A})$.
- Si a es un vector de dimensión $p \times 1$, entonces $a'a = tr(aa')$.

1.2. Determinantes

El determinante de una matriz $\boldsymbol{A}$, se define como,

$$det(\boldsymbol{A}) = \begin{cases} \sum_{\sigma \in S_m} \left(sgn(\sigma) \prod_{i=1}^{m} a_{i,\sigma_i} \right) & \text{para } m = p, \\ \text{no definido} & \text{para } m \neq p. \end{cases}$$

1.2. Determinantes

También puede definirse como,

$$det(\boldsymbol{A}) = \begin{cases} \sum_{i=1}^{m}(-1)^{j+1}a_{1j}det(\boldsymbol{A}_{1j}) & \text{para } m = p, \\ \text{no definido} & \text{para } m \neq p, \end{cases}$$

donde $\boldsymbol{A}_{1j}$ es la matriz reducida (o submatriz) de $\boldsymbol{A}$ eliminando el renglón 1 y la columna j. El producto

$$(-1)^{j+1}det(\boldsymbol{A}_{1j}),$$

se conoce como cofactor j-ésimo de $\boldsymbol{A}$.

1.2. Determinantes

El operador $det(\cdot)$ tiene varias propiedades interesantes, entre ellas:

- $det(c\boldsymbol{A}) = c^m \cdot det(\boldsymbol{A})$ para todo escalar c.
- $det(\boldsymbol{A}) = det(\boldsymbol{A}')$.
- $det(\boldsymbol{AB}) = det(\boldsymbol{A})det(\boldsymbol{B})$.
- $det(\boldsymbol{A}^m) = det(\boldsymbol{A})^m$.

2. Matrices inversas

La inversa de una matriz $\boldsymbol{A}$ se define como,

$$\boldsymbol{A}^{-1} := \begin{cases} \boldsymbol{A}^{-1} & \text{para } m = p \text{ y } \boldsymbol{A} \text{ invertible,} \\ \text{no definida} & \text{para } m \neq p \text{ o } \boldsymbol{A} \text{ no invertible.} \end{cases}$$

Se dice que $\boldsymbol{A}$ es invertible (no singular o no degenerada) si existe $\boldsymbol{A}^{-1}$ tal que

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I},$$

donde $\boldsymbol{I}$ es una matriz identidad de orden $m \times m$, i.e.

$$\boldsymbol{I} = (\iota_{ij})_{i=1,j=1}^{m,m},$$

con

$$\iota_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

2.1. Propiedades básicas

Las matrices inversas, cuando existen, cumplen con varias propiedades importantes. Entre ellas:

- $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.
- $(c\boldsymbol{A})^{-1} = \frac{1}{c}\boldsymbol{A}^{-1}$.
- $det(\boldsymbol{A}^{-1}) = \frac{1}{det(\boldsymbol{A})}$.

Cuando la matriz inversa existe, se puede calcular como

$$\boldsymbol{A}^{-1} = \frac{adj(\boldsymbol{A})}{det(\boldsymbol{A})},$$

si $det(\boldsymbol{A}) \neq 0$, donde $adj(\boldsymbol{A})$ denota la matriz adjunta de $\boldsymbol{A}$, definida como la matriz transpuesta de los cofactores, i.e.

$$adj(\boldsymbol{A}) = (cof(A, j, k))',$$

donde

$$cof(A, j, k) = (-1)^{j+k} det(\boldsymbol{A}_{jk}),$$

siendo $\boldsymbol{A}_{jk}$ la matriz reducida de $\boldsymbol{A}$.
La matriz inversa $\boldsymbol{A}^{-1}$ cuando existe es única si $det(\boldsymbol{A}) > 0$.

2.2. Inversas generalizadas

Cuando $det(\boldsymbol{A})$ no es invertible de manera única, es posible hacer noción a la matriz inversa generalizada, definida como la matriz $\boldsymbol{A}^{-1}$ tal que

$$\boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{A}.$$

La matriz inversa generalizada no es única. El número de inversas generalizadas para $\boldsymbol{A}$ estará en función de su *rango*.
El rango de la matriz $\boldsymbol{A}$, $rank(\boldsymbol{A})$, se define como el número máximo de vectores en $\boldsymbol{A}$ (renglones o columnas) que son linealmente independientes.
Se dice que una matriz es de *rango completo* si $rank(\boldsymbol{A}) = \min\{m, p\}$.
Si el rango de una matriz cuadrada $\boldsymbol{A}$ es

$$rank(\boldsymbol{A}) = r < m,$$

entonces el número de matrices inversas generalizadas admisibles para $\boldsymbol{A}$ es $m - r + 1$.

3.1. Solución de sistemas lineales

Consideremos el sistema lineal para $\boldsymbol{A}$, de la forma,

$$\boldsymbol{A}x = b,$$

donde x es un vector de dimensión $p \times 1$ y b es un vector de dimensión $m \times 1$. A partir de este sistema, podemos definir la matriz aumentada $\boldsymbol{B}$ como,

$$\boldsymbol{B} = [\boldsymbol{A}\ b].$$

La existencia de x (solución del sistema lineal) estará determinada en términos de que se satisfagan ciertas propiedades de $\boldsymbol{A}$ y $\boldsymbol{B}$ conjuntamente:

- Si $rank(\boldsymbol{A}) = rank(\boldsymbol{B}) = m$ se inducirá una solución única para x.
- Si $rank(\boldsymbol{A}) = rank(\boldsymbol{B}) = r < m$ se inducirán $(m - r + 1)$ soluciones distintas para x.
- Si $rank(\boldsymbol{A}) < rank(\boldsymbol{B})$ se inducirá que no exista solución para x.

En el caso en que $\boldsymbol{A}$ sea cuadrada e invertible, la solución para x es

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

3.2. Eigenvalores y eigenvectores

Los eigenvalores, λ_i's, y los eigenvectores, v_i's, de una matriz $\boldsymbol{A}$ son tales que satisfacen la siguiente relación

$$\boldsymbol{A} v_i = \lambda_i v_i,$$

donde λ_i es un escalar y v_i es un vector de dimensión $m \times 1$.
Definiendo la matriz $\boldsymbol{V}$ con columnas dadas por los vectores v_i se tiene que

$$\boldsymbol{A}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{D},$$

con $\boldsymbol{D} = (\delta_{ij}\lambda_i)$

3.2. Eigenvalores y eigenvectores

Existen propiedades importantes de eigenvectores y eigenvalores:

- $eig(\boldsymbol{AB}) = eig(\boldsymbol{BA})$.
- $rank(\boldsymbol{A}) = r \leq m$ entonces existen r egivenvalores λ_i distintos de cero.
- Si $\boldsymbol{A}$ es simétrica, entonces $\boldsymbol{AA}' = \boldsymbol{I}$, en cuyo caso:
    - $\boldsymbol{A}$ es ortogonal,
    - λ_i son reales,
    - $eig(\boldsymbol{A}^{-1})$ son λ_i^{-1}s.

3.3. Descomposición SVD

La descomposición de valores singulares de una matriz $\boldsymbol{A}$ de dimensión $m \times p$ puede escribirse como:

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{-1},$$

donde

- $\boldsymbol{U}$ son los eigenvectores de la matriz $\boldsymbol{A}\boldsymbol{A}'$
- $\boldsymbol{D}$ es la matriz diagonal con los eigenvalores de $\boldsymbol{A}\boldsymbol{A}'$
- $\boldsymbol{V}$ es la matriz de eigenvectores de la matriz $\boldsymbol{A}'\boldsymbol{A}$.

4. Ilustración

Revisemos los datos de `R-base` que hacen referencia a las provincias en Suiza:

```
data(swiss)
?swiss
dimnames(swiss)
```

```
## [[1]]
##  [1] "Courtelary"   "Delemont"     "Franches-Mnt" "Moutier"
##  [5] "Neuveville"   "Porrentruy"   "Broye"        "Glane"
##  [9] "Gruyere"      "Sarine"       "Veveyse"      "Aigle"
## [13] "Aubonne"      "Avenches"     "Cossonay"     "Echallens"
## [17] "Grandson"     "Lausanne"     "La Vallee"    "Lavaux"
## [21] "Morges"       "Moudon"       "Nyone"        "Orbe"
## [25] "Oron"         "Payerne"      "Paysd'enhaut" "Rolle"
## [29] "Vevey"        "Yverdon"      "Conthey"      "Entremont"
## [33] "Herens"       "Martigwy"     "Monthey"      "St Maurice"
## [37] "Sierre"       "Sion"         "Boudry"       "La Chauxdfnd"
## [41] "Le Locle"     "Neuchatel"    "Val de Ruz"   "ValdeTravers"
## [45] "V. De Geneve" "Rive Droite"  "Rive Gauche"
```

4.1. Desomposición cruda

Realizamos la descomposición SVD de la matriz $\boldsymbol{Y}$ de 47×6:

```
udv<-svd(Y)
summary(udv)
```

```
##   Length Class  Mode
## d   6    -none- numeric
## u 282    -none- numeric
## v  36    -none- numeric
```

```
plot(udv$d,type="h",lwd=4,col="gray")
```

4.2. Descomposición estandarizada por columnas

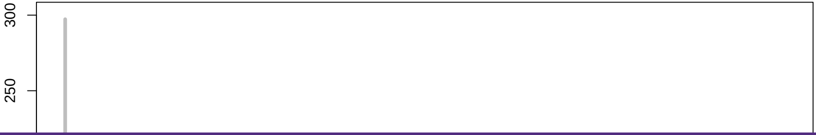
Calculamos la descomposición con datos estandarizados (por columnas):

- Estandarización:

```
m<-nrow(Y)
Im<-diag(m)
Om<-rep(1,m)
Ycc<-(Im - Om%*%t(Om)/m) %*% Y
```

- Descomposición SVD:

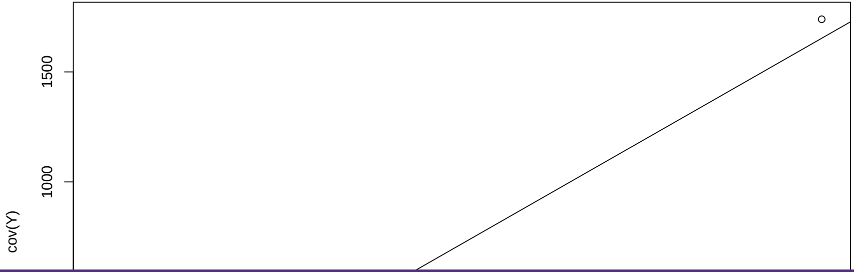
```
udvcc<-svd(Ycc)

plot(udvcc$d,type="h",lwd=4,col="gray")
```



4.3. Descomposición estandarizada general

- Calculamos aproximaciones a la matriz de covarianzas:

```
for(k in 1:3)
{
  K<-seq(1,k,length=k)
  VDV<- udvcc$v[,K,drop=FALSE] %*% diag(udvcc$d^2)[K,K] %*%
     t( udvcc$v[,K,drop=FALSE] )
  plot(VDV/m, cov(Y)) ; abline(0,1)
}
```



5. Ejercicios

1. Estudien la descomposición de Cholesky.
2. Calculen la descomposición SVD con otras matrices diferentes a la de `swiss`.
3. Comparen los resultados de la descomposición SVD de los datos `swiss` con los datos originales, estandarizados por columnas y estandarizados general.

Referencia

- Press (2005) *Applied Multivariate Analysis*, capítulo 2.

Referencia extendida

- Gentle, James E. (2007) *Matrix Algebra: Theory, Computations, and Applications in Statistics*. London: Springer

Table of Contents