Estadística Multivariada Computacional
Matriz de Datos

Mathias Bourel

IMERL - Facultad de Ingeniería, Universidad de la República, Uruguay

27 de marzo de 2019

Plan

Plan

Consideramos bases de datos del tipo

$$X|Y$$

con $X \in \mathcal{M}_{n\times p}$ e $Y \in \mathcal{M}_{n\times 1}$ (etiqueta).
Podemos pensar las p columnas de X como realización de variables aleatorias reales o categóricas y las n filas son los valores que toma cada individuo en cada una de estas variables.

Usaremos este tipo de bases de datos para hacer inferencia y construir un predictor f que dada una nueva observación pueda predecir una categoría o un valor habiendo aprendido de las observaciones de la base de datos.

Introducción

Consideramos bases de datos del tipo

$$X|Y$$

con $X \in \mathcal{M}_{n\times p}$ e $Y \in \mathcal{M}_{n\times 1}$ (etiqueta).
Podemos pensar las p columnas de X como realización de variables aleatorias reales o categóricas y las n filas son los valores que toma cada individuo en cada una de estas variables.

Usaremos este tipo de bases de datos para hacer inferencia y construir un predictor f que dada una nueva observación pueda predecir una categoría o un valor habiendo aprendido de las observaciones de la base de datos.

Es decir buscamos una función f que logra predecir y en función de las p variables $x_1, \dots, x_p$, i.e:

$$\widehat{y} = f(x_1, \dots, x_d)$$

Consideramos bases de datos del tipo

$$X|Y$$

con $X \in \mathcal{M}_{n\times p}$ e $Y \in \mathcal{M}_{n\times 1}$ (etiqueta).
Podemos pensar las p columnas de X como realización de variables aleatorias reales o categóricas y las n filas son los valores que toma cada individuo en cada una de estas variables.

Usaremos este tipo de bases de datos para hacer inferencia y construir un predictor f que dada una nueva observación pueda predecir una categoría o un valor habiendo aprendido de las observaciones de la base de datos.

Es decir buscamos una función f que logra predecir y en función de las p variables $x_1, \ldots, x_p$, i.e:

$$\widehat{y} = f(x_1, \ldots, x_d)$$

Por ejemplo en los datos Advertising:

```
> datos=read.csv("Advertising.csv",header=T,sep=",")
> datos
      X    TV Radio Newspaper Sales
1     1 230.1  37.8      69.2  22.1
2     2  44.5  39.3      45.1  10.4
3     3  17.2  45.9      69.3   9.3
4     4 151.5  41.3      58.5  18.5
5     5 180.8  10.8      58.4  12.9
```

Por ejemplo en los datos Advertising:

```
> datos=read.csv("Advertising.csv",header=T,sep=",")
> datos
      X    TV Radio Newspaper Sales
1     1 230.1  37.8      69.2  22.1
2     2  44.5  39.3      45.1  10.4
3     3  17.2  45.9      69.3   9.3
4     4 151.5  41.3      58.5  18.5
5     5 180.8  10.8      58.4  12.9
```

En este caso:

- $X = (X_1, X_2, X_3)$ es el vector de entradas, *input*, de variables independientes:
  - X_1 es el presupuesto asignado a TV
  - X_2 es el presupuesto asignado a Radio
  - X_3 es el presupuesto asignado a Newspaper
- Y es el monto de las ventas realizadas y es la variable de salida, de respuesta, el *output*, o variable dependiente.

Introducción

Por ejemplo en los datos Advertising:

```
> datos=read.csv("Advertising.csv",header=T,sep=",")
> datos
      X    TV Radio Newspaper Sales
1     1 230.1  37.8      69.2  22.1
2     2  44.5  39.3      45.1  10.4
3     3  17.2  45.9      69.3   9.3
4     4 151.5  41.3      58.5  18.5
5     5 180.8  10.8      58.4  12.9
```

En este caso:

- $X = (X_1, X_2, X_3)$ es el vector de entradas, *input*, de variables independientes:
    - X_1 es el presupuesto asignado a TV
    - X_2 es el presupuesto asignado a Radio
    - X_3 es el presupuesto asignado a Newspaper
- Y es el monto de las ventas realizadas y es la variable de salida, de respuesta, el *output*, o variable dependiente.

En general vamos a querer modelos de la forma general:

$$Y = f(X_1, \ldots, X_p) + \epsilon$$

donde $X_1, X_2, \ldots, X_p$ son variables predictoras e Y es la variable de respuesta, ϵ es el término de error, independiente de X y con media 0.

Introducción

En el contexto de la inferencia nos hacemos a menudo las preguntas:

1. De todas las variables de entradas, ¿cuales son los predictores que están asociados a la variable de respuesta?
2. ¿Cual es la relación entre cada predictor y la variable de respuesta?
3. ¿Existe una relación lineal, fácil de explicar, entre los predictores y la variable de respuesta, o esta relación es más complicada?

Introducción

En el contexto de la inferencia nos hacemos a menudo las preguntas:

1. De todas las variables de entradas, ¿cuales son los predictores que están asociados a la variable de respuesta?
2. ¿Cual es la relación entre cada predictor y la variable de respuesta?
3. ¿Existe una relación lineal, fácil de explicar, entre los predictores y la variable de respuesta, o esta relación es más complicada?

En este curso veremos numerosos ejemplos que combinan estos dos enfoques (predicción e inferencia).

En el contexto de la inferencia nos hacemos a menudo las preguntas:

1. De todas las variables de entradas, ¿cuales son los predictores que están asociados a la variable de respuesta?
2. ¿Cual es la relación entre cada predictor y la variable de respuesta?
3. ¿Existe una relación lineal, fácil de explicar, entre los predictores y la variable de respuesta, o esta relación es más complicada?

En este curso veremos numerosos ejemplos que combinan estos dos enfoques (predicción e inferencia).

Dependiendo del objetivo que se tenga, distintos métodos para estimar f serán usados. Por ejemplo, los modelos lineales tienen interpretaciones claras en cuanto a la inferencia pero muchas veces, y dependiendo de los datos, poco poder predictivo, por lo que se usará o complementará con un modelo con mejor performance, pero quizás sea menos interpretable.

Introducción

Hay dos maneras de ver la matriz de datos.
$\mathbf{X} = ((x_{ij}))_{i=1,\ldots,n}^{j=1,\ldots,p} \in \mathcal{M}_{n\times p}$ matriz de datos (n datos con p variables).

Introducción

Hay dos maneras de ver la matriz de datos.
$\mathbf{X} = ((x_{ij}))_{i=1,\dots,n}^{j=1,\dots,p} \in \mathcal{M}_{n\times p}$ matriz de datos (n datos con p variables).

Por filas (individuos):

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{11} & x_{12} & \dots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} = \begin{pmatrix} \text{individuo 1} \\ \text{individuo 2} \\ \vdots \\ \text{individuo n} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{pmatrix}$$

Hay dos maneras de ver la matriz de datos.
$\mathbf{X} = ((x_{ij}))_{i=1,\dots,n}^{j=1,\dots,p} \in \mathcal{M}_{n\times p}$ matriz de datos (n datos con p variables).

Por filas (individuos):

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{11} & x_{12} & \dots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} = \begin{pmatrix} \text{individuo 1} \\ \text{individuo 2} \\ \vdots \\ \text{individuo n} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{pmatrix}$$

Por columnas (características):

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{11} & x_{12} & \dots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} = \begin{pmatrix} v & v & & v \\ a & a & & a \\ r & r & & r \\ i & i & \dots & i \\ a & a & & a \\ b & b & & b \\ l & l & & l \\ e & e & & e \\ 1 & 2 & & p \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & \dots & x_p \end{pmatrix}$$

Hay dos maneras de ver la matriz de datos.
$\mathbf{X} = ((x_{ij}))_{i=1,\ldots,n}^{j=1,\ldots,p} \in \mathcal{M}_{n\times p}$ matriz de datos (n datos con p variables).

Por filas (individuos):

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1p} \\ x_{11} & x_{12} & \ldots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{np} \end{pmatrix} = \begin{pmatrix} \text{individuo 1} \\ \text{individuo 2} \\ \vdots \\ \text{individuo n} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{pmatrix}$$

Por columnas (características):

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1p} \\ x_{11} & x_{12} & \ldots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{np} \end{pmatrix} = \begin{pmatrix} v & v & & v \\ a & a & & a \\ r & r & & r \\ i & i & \ldots & i \\ a & a & & a \\ b & b & & b \\ l & l & & l \\ e & e & & e \\ 1 & 2 & & p \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & \ldots & x_p \end{pmatrix}$$

Notamos por y_i al valor de respuesta del individuo i. Nuestro set de datos es entonces:

$$\mathcal{L} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)\}$$

Plan

- Una matriz con n filas y p columnas se dice que tiene tamaño $n \times p$:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \in \mathcal{M}_{n\times p}$$

Notamos por x_{ij} el elemento que está en la intersección de la fila i con la columna j.

- Una matriz con n filas y p columnas se dice que tiene tamaño $n \times p$:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \in \mathcal{M}_{n \times p}$$

Notamos por x_{ij} el elemento que está en la intersección de la fila i con la columna j.

- La traspuesta de una matriz $X \in \mathcal{M}_{n \times p}$ es la matriz $X' \in \mathcal{M}_{p \times n}$ definida por $x_{ij}^{'} = x_{ji}$ para todo i, j.

Por ejemplo, si $X = \begin{pmatrix} 2 & -1 & 3 & 0 \\ 1 & 2 & -3 & 2 \\ 6 & 3 & -2 & 10 \end{pmatrix}$ entonces $X' = \begin{pmatrix} 2 & 1 & 6 \\ -1 & 2 & 3 \\ 3 & -3 & 2 \\ 0 & 2 & 10 \end{pmatrix}$

- Una matriz con n filas y p columnas se dice que tiene tamaño $n \times p$:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \in \mathcal{M}_{n \times p}$$

Notamos por x_{ij} el elemento que está en la intersección de la fila i con la columna j.

- La traspuesta de una matriz $X \in \mathcal{M}_{n \times p}$ es la matriz $X' \in \mathcal{M}_{p \times n}$ definida por $x'_{ij} = x_{ji}$ para todo i, j.

Por ejemplo, si $X = \begin{pmatrix} 2 & -1 & 3 & 0 \\ 1 & 2 & -3 & 2 \\ 6 & 3 & -2 & 10 \end{pmatrix}$ entonces $X' = \begin{pmatrix} 2 & 1 & 6 \\ -1 & 2 & 3 \\ 3 & -3 & 2 \\ 0 & 2 & 10 \end{pmatrix}$

- Un vector columna $\mathbf{x}$ es un matriz con una única columna y un vector fila $\mathbf{x}'$ es una matriz con una sola fila. Si $\mathbf{x} = \begin{pmatrix} 1 \\ 4 \\ -3 \end{pmatrix}$ entonces $\mathbf{x}' = \begin{pmatrix} 1 & 4 & -3 \end{pmatrix}$.

- **Suma de matrices**. Si X e Y son dos matrices $n \times p$, entonces sumamos entrada a entrada, obteniendo una nueva matriz de tamaño $n \times p$. Por ejemplo, si $X = \begin{pmatrix} 2 & -1 & 3 & 0 \\ 1 & 2 & -3 & 2 \\ 6 & 3 & -2 & 10 \end{pmatrix}$ e $Y = \begin{pmatrix} 4 & 2 & 5 & -1 \\ 8 & 1 & 3 & 4 \\ 2 & -3 & 1 & -3 \end{pmatrix}$ entonces

$$X + Y = \begin{pmatrix} 6 & 1 & 8 & -1 \\ 9 & 3 & 0 & 6 \\ 8 & 0 & -1 & 7 \end{pmatrix}$$

- **Multiplicación por un escalar**. Si X es de tamaño $n \times p$ y $\lambda \in \mathbb{R}$ entonces la matriz λX es la matriz que se obtiene de X multiplicando todas sus entradas por λ. Por ejemplo si $X = \begin{pmatrix} 2 & -1 & 3 & 0 \\ 1 & 2 & -3 & 2 \\ 6 & 3 & -2 & 10 \end{pmatrix}$ entonces

$$3X = \begin{pmatrix} 6 & -3 & 9 & 0 \\ 3 & 6 & -9 & 6 \\ 18 & 9 & -6 & 30 \end{pmatrix}$$

Multiplicación de matrices

Si $X = ((x_{ij})) \in \mathcal{M}_{n\times p}$ e $Y = ((y_{ij})) \in \mathcal{M}_{p\times m}$ entonces la matriz $Z = XY \in \mathcal{M}_{n\times m}$ y la entrada ij de Z es

$$z_{ij} = \sum_{k=1}^{p} x_{ik} y_{kj}$$

Por ejemplo si $X = \begin{pmatrix} 2 & 3 \\ -1 & 4 \end{pmatrix}$ e $Y = \begin{pmatrix} -1 & 2 & -2 \\ 3 & 2 & -4 \end{pmatrix}$ entonces

$$Z = XY = \begin{pmatrix} 7 & 10 & -14 \\ 13 & 6 & -14 \end{pmatrix}$$

Cuidado con el orden:

- Observar que en el ejemplo anterior no se puede hacer YX...
- Si $\mathbf{u} \in \mathbb{R}^n$ es un vector entonces $\mathbf{u}'\mathbf{u} \in \mathbb{R}$ pero $\mathbf{u}\mathbf{u}' \in \mathcal{M}_{n\times n}$...

La traspuesta de una matriz verifica:

- $(X + Y)' = X' + Y'$
- $(aX)' = aX'$
- $(XY)' = Y'X'$

Combinación lineal

- Si $\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$ y $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}$ entonces

$$\mathbf{X}\beta = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \cdot \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} = \beta_1 \begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{pmatrix} + \beta_2 \begin{pmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{n2} \end{pmatrix} + \cdots + \beta_p \begin{pmatrix} x_{1p} \\ x_{2p} \\ \vdots \\ x_{np} \end{pmatrix}$$

  es una combinación lineal de las columnas de $\mathbf{X}$.

- Si $\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$ entonces el núcleo de $\mathbf{X}$ es

$$N(\mathbf{X}) = \{\beta \in \mathbb{R}^p : \mathbf{X}\beta = \mathbf{0}\}$$

  lo cual implica que si $\beta \in N(\mathbf{X})$ y es no nulo entonces una columna de $\mathbf{X}$ es combinación lineal de las demás.

Un vector fila $\mathbf{x}' = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix}$ o un vector columna $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ es un punto en $\mathbb{R}^n$.

El *producto escalar* entre dos vectores $\mathbf{x}$ e $\mathbf{y}$ de $\mathbb{R}^n$ se nota por

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i = \mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x}$$

Se dice que dos vectores son *ortogonales* si $\langle \mathbf{x}, \mathbf{y} \rangle = 0$
La *norma* del vector $\mathbf{x}$ es

$$||x|| = \sqrt{x_1^2 + x_2^2 + \dots x_n^2} = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$$

Es posible probar que

$$\langle \mathbf{x}, \mathbf{y} \rangle = ||\mathbf{x}|| ||\mathbf{y}|| cos(\mathbf{x}, \mathbf{y})$$

Proyección ortogonal

Sea **u** un vector no nulo. Entonces la *proyección ortogonal* de **x** sobre **u** es el vector

$$P_{\mathbf{u}}(\mathbf{x}) = \frac{\langle \mathbf{x}, \mathbf{u} \rangle}{||\mathbf{u}||^2}\mathbf{u} = \frac{\langle \mathbf{x}, \mathbf{u} \rangle}{||\mathbf{u}||}\frac{\mathbf{u}}{||\mathbf{u}||}$$

Observar que si **u** es unitario (tiene norma 1) entonces

$$P_{\mathbf{u}}(\mathbf{x}) = (\mathbf{x}'\mathbf{u})\mathbf{u}$$

La coordenada de $P_{\mathbf{u}}(\mathbf{x})$ sobre **u** es $\frac{\langle \mathbf{x}, \mathbf{u} \rangle}{||\mathbf{u}||}$.

Por ejemplo si $\mathbf{x} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$ y $\mathbf{u} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ entonces la proyección ortogonal de **x** sobre **u** es:

$$P_{\mathbf{u}}(\mathbf{x}) = \frac{1}{5}\begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

y la coordenada de $P_{\mathbf{u}}(\mathbf{x})$ es $\frac{1}{\sqrt{5}}$.

Proyección ortogonal

Sea **u** un vector no nulo. Entonces la *proyección ortogonal* de **x** sobre **u** es el vector

$$P_{\mathbf{u}}(\mathbf{x}) = \frac{\langle \mathbf{x}, \mathbf{u} \rangle}{||\mathbf{u}||^2}\mathbf{u} = \frac{\langle \mathbf{x}, \mathbf{u} \rangle}{||\mathbf{u}||}\frac{\mathbf{u}}{||\mathbf{u}||}$$

Observar que si **u** es unitario (tiene norma 1) entonces

$$P_{\mathbf{u}}(\mathbf{x}) = (\mathbf{x}'\mathbf{u})\mathbf{u}$$

La coordenada de $P_{\mathbf{u}}(\mathbf{x})$ sobre **u** es $\frac{\langle \mathbf{x}, \mathbf{u} \rangle}{||\mathbf{u}||}$.
Por ejemplo si $\mathbf{x} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$ y $\mathbf{u} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ entonces la proyección ortogonal de **x** sobre **u** es:

$$P_{\mathbf{u}}(\mathbf{x}) = \frac{1}{5}\begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

y la coordenada de $P_{\mathbf{u}}(\mathbf{x})$ es $\frac{1}{\sqrt{5}}$.

Si $\{\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_r}\}$ es una base ortogonal de S, entonces la proyección ortogonal de **x** sobre S es el vector

$$P_S(x) = \frac{\mathbf{x}'\mathbf{u_1}}{||\mathbf{u_1}||^2}\mathbf{u_1} + \frac{\mathbf{x}'\mathbf{u_2}}{||\mathbf{u_2}||^2}\mathbf{u_2} + \cdots + \frac{\mathbf{x}'\mathbf{u_r}}{||\mathbf{u_r}||^2}\mathbf{u_r}$$

Sea $A \in \mathcal{M}_{n \times n}$. Decimos que $\lambda \in \mathbb{R}$ es un *valor propio* de A si existe un vector $\mathbf{x} \in \mathbb{R}^n$ no nulo tal que

$$A\mathbf{x} = \lambda \mathbf{x}$$

Al vector $\mathbf{x}$ se le llama *vector propio* asociado al valor propio λ.

Sea $A \in \mathcal{M}_{n\times n}$. Decimos que $\lambda \in \mathbb{R}$ es un *valor propio* de A si existe un vector $\mathbf{x} \in \mathbb{R}^n$ no nulo tal que

$$A\mathbf{x} = \lambda\mathbf{x}$$

Al vector $\mathbf{x}$ se le llama *vector propio* asociado al valor propio λ.

Una matriz $A \in \mathcal{M}_{n\times n}$ es *simétrica* si $A' = A$.

Algunas propiedades matriciales

Sea $A \in \mathcal{M}_{n \times n}$. Decimos que $\lambda \in \mathbb{R}$ es un *valor propio* de A si existe un vector $\mathbf{x} \in \mathbb{R}^n$ no nulo tal que

$$A\mathbf{x} = \lambda \mathbf{x}$$

Al vector $\mathbf{x}$ se le llama *vector propio* asociado al valor propio λ.

Una matriz $A \in \mathcal{M}_{n \times n}$ es *simétrica* si $A' = A$.

Una matriz $A \in \mathcal{M}_{n \times n}$ es *invertible* si exite una matriz $B \in \mathcal{M}_{n \times n}$ tal que $AB = BA = I_n$.

Sea $A \in \mathcal{M}_{n\times n}$. Decimos que $\lambda \in \mathbb{R}$ es un *valor propio* de A si existe un vector $\mathbf{x} \in \mathbb{R}^n$ no nulo tal que

$$A\mathbf{x} = \lambda\mathbf{x}$$

Al vector $\mathbf{x}$ se le llama *vector propio* asociado al valor propio λ.

Una matriz $A \in \mathcal{M}_{n\times n}$ es *simétrica* si $A' = A$.

Una matriz $A \in \mathcal{M}_{n\times n}$ es *invertible* si exite una matriz $B \in \mathcal{M}_{n\times n}$ tal que $AB = BA = I_n$.
Si A y B son invertibles entonces AB es invertible y $(AB)^{-1} = B^{-1}A^{-1}$.

Sea $A \in \mathcal{M}_{n\times n}$. Decimos que $\lambda \in \mathbb{R}$ es un *valor propio* de A si existe un vector $\mathbf{x} \in \mathbb{R}^n$ no nulo tal que

$$A\mathbf{x} = \lambda \mathbf{x}$$

Al vector $\mathbf{x}$ se le llama *vector propio* asociado al valor propio λ.

Una matriz $A \in \mathcal{M}_{n\times n}$ es *simétrica* si $A' = A$.

Una matriz $A \in \mathcal{M}_{n\times n}$ es *invertible* si exite una matriz $B \in \mathcal{M}_{n\times n}$ tal que $AB = BA = I_n$.
Si A y B son invertibles entonces AB es invertible y $(AB)^{-1} = B^{-1}A^{-1}$.

Una matriz $U \in \mathcal{M}_{n\times n}$ es *ortogonal* si U es invertible y $U'U = UU' = I_n$.

Algunas propiedades matriciales

- Si $A \in \mathcal{M}_{n\times n}$ es una matriz simétrica entonces los vectores propios asociados a valores propios distintos son ortogonales.

Si $A \in \mathcal{M}_{n\times p}$ es una matriz entonces:

- $AA' \in \mathcal{M}_{n\times n}$ y $A'A \in \mathcal{M}_{p\times p}$ son simétricas.
- $rg(A'A) = rg(AA') = rg(A)$.
- Los valores propios de $A'A$ son no negativos.
- Los valores propios no nulos de AA' y de $A'A$ coinciden.

Si $A \in \mathcal{M}_{n \times n}$ es una matriz simétrica entonces los vectores propios asociados a valores propios distintos son ortogonales.

Si $A \in \mathcal{M}_{n\times n}$ es una matriz simétrica entonces los vectores propios asociados a valores propios distintos son ortogonales.

Para ver esto:

1. Si $A \in \mathcal{M}_{n\times n}$ es una matriz cualquiera entonces

$$\langle Ax, y\rangle = \langle x, A'y\rangle$$

   Esto es porque $\langle x, y\rangle = x'y$

Si $A \in \mathcal{M}_{n \times n}$ es una matriz simétrica entonces los vectores propios asociados a valores propios distintos son ortogonales.

Para ver esto:

1. Si $A \in \mathcal{M}_{n \times n}$ es una matriz cualquiera entonces
$$\langle Ax, y \rangle = \langle x, A'y \rangle$$
Esto es porque $\langle x, y \rangle = x'y$
2. Si $A \in \mathcal{M}_{n \times n}$ es una matriz simétrica entonces
$$\langle Ax, y \rangle = \langle x, Ay \rangle$$

Algunas propiedades matriciales

Si $A \in \mathcal{M}_{n\times n}$ es una matriz simétrica entonces los vectores propios asociados a valores propios distintos son ortogonales.

Para ver esto:

1. Si $A \in \mathcal{M}_{n\times n}$ es una matriz cualquiera entonces
$$\langle Ax, y\rangle = \langle x, A'y\rangle$$
Esto es porque $\langle x, y\rangle = x'y$
2. Si $A \in \mathcal{M}_{n\times n}$ es una matriz simétrica entonces
$$\langle Ax, y\rangle = \langle x, Ay\rangle$$
3. Sea A simétrica ahora y consideramos u_1 vector propio de A asociado a λ_1 y u_2 vector propio de A asociado a λ_2 entonces
$$\langle Au_1, u_2\rangle = \langle \lambda_1 u_1, u_2\rangle = \lambda_1\langle u_1, u_2\rangle$$
que es igual a
$$\langle u_1, Au_2\rangle = \langle u_1, \lambda_2 u_2\rangle = \lambda_2\langle u_1, u_2\rangle$$
Por lo tanto
$$(\lambda_1 - \lambda_2)\langle u_1, u_2\rangle = 0$$
y al ser $\lambda_1 \neq \lambda_2$ se tiene que $u_1 \perp u_2$.

Algunas propiedades matriciales

Si $A \in \mathcal{M}_{n\times p}$ es una matriz entonces:

- $AA' \in \mathcal{M}_{n\times n}$ y $A'A \in \mathcal{M}_{p\times p}$ son simétricas.

Algunas propiedades matriciales

Si $A \in \mathcal{M}_{n\times p}$ es una matriz entonces:

- $AA' \in \mathcal{M}_{n\times n}$ y $A'A \in \mathcal{M}_{p\times p}$ son simétricas.
- $rg(A'A) = rg(AA') = rg(A)$.

Para probar que $rg(A'A) = rg(A)$ probamos que $N(A'A) = N(A)$ ya que $nu(B) + rg(B) = p$ siendo p la cantidad de columnas de una matriz B y $nu(B) = \dim N(B)$. Es claro que $N(A) \subset N(A'A)$ para probar la otra inclusión, si $x \in N(A'A)$ entonces $A'Ax = 0$ y entonces

$$0 = \langle A'Ax, x\rangle = \langle Ax, Ax\rangle = ||Ax||^2 \Rightarrow Ax = 0$$

por lo que $x \in N(A)$.

Algunas propiedades matriciales

Si $A \in \mathcal{M}_{n\times p}$ es una matriz entonces:

- $AA' \in \mathcal{M}_{n\times n}$ y $A'A \in \mathcal{M}_{p\times p}$ son simétricas.
- $rg(A'A) = rg(AA') = rg(A)$.

Para probar que $rg(A'A) = rg(A)$ probamos que $N(A'A) = N(A)$ ya que $nu(B) + rg(B) = p$ siendo p la cantidad de columnas de una matriz B y $nu(B) = \dim N(B)$. Es claro que $N(A) \subset N(A'A)$ para probar la otra inclusión, si $x \in N(A'A)$ entonces $A'Ax = 0$ y entonces

$$0 = \langle A'Ax, x\rangle = \langle Ax, Ax\rangle = ||Ax||^2 \Rightarrow Ax = 0$$

por lo que $x \in N(A)$. Para probar que $rg(AA') = rg(A)$ usamos que $rg(A') = rg(A)$ y hacemos un razonamiento análogo al anterior.

Algunas propiedades matriciales

Si $A \in \mathcal{M}_{n\times p}$ es una matriz entonces:

- $AA' \in \mathcal{M}_{n\times n}$ y $A'A \in \mathcal{M}_{p\times p}$ son simétricas.

- $rg(A'A) = rg(AA') = rg(A)$.

Para probar que $rg(A'A) = rg(A)$ probamos que $N(A'A) = N(A)$ ya que $nu(B) + rg(B) = p$ siendo p la cantidad de columnas de una matriz B y $nu(B) = \dim N(B)$. Es claro que $N(A) \subset N(A'A)$ para probar la otra inclusión, si $x \in N(A'A)$ entonces $A'Ax = 0$ y entonces

$$0 = \langle A'Ax, x\rangle = \langle Ax, Ax\rangle = ||Ax||^2 \Rightarrow Ax = 0$$

por lo que $x \in N(A)$. Para probar que $rg(AA') = rg(A)$ usamos que $rg(A') = rg(A)$ y hacemos un razonamiento análogo al anterior.

- Los valores propios de $A'A$ son no negativos.

En efecto si λ es valor propio $\langle A'Ax, x\rangle = \lambda||x||^2$ por un lado, y $\langle A'Ax, x\rangle = \langle Ax, Ax\rangle = ||Ax||^2 \geq 0$ por otro. Entonces

$$\lambda||x||^2 = ||Ax||^2 \geq 0 \Rightarrow \lambda = \frac{||Ax||^2}{||x||^2}$$

ya que $||x||^2 \neq 0$ por ser x vector propio y por lo tanto $\lambda \geq 0$.

Algunas propiedades matriciales

Si $A \in \mathcal{M}_{n\times p}$ es una matriz entonces:

- $AA' \in \mathcal{M}_{n\times n}$ y $A'A \in \mathcal{M}_{p\times p}$ son simétricas.

- $rg(A'A) = rg(AA') = rg(A)$.

Para probar que $rg(A'A) = rg(A)$ probamos que $N(A'A) = N(A)$ ya que $nu(B) + rg(B) = p$ siendo p la cantidad de columnas de una matriz B y $nu(B) = \dim N(B)$. Es claro que $N(A) \subset N(A'A)$ para probar la otra inclusión, si $x \in N(A'A)$ entonces $A'Ax = 0$ y entonces

$$0 = \langle A'Ax, x\rangle = \langle Ax, Ax\rangle = ||Ax||^2 \Rightarrow Ax = 0$$

por lo que $x \in N(A)$. Para probar que $rg(AA') = rg(A)$ usamos que $rg(A') = rg(A)$ y hacemos un razonamiento análogo al anterior.

- Los valores propios de $A'A$ son no negativos.

En efecto si λ es valor propio $\langle A'Ax, x\rangle = \lambda||x||^2$ por un lado, y $\langle A'Ax, x\rangle = \langle Ax, Ax\rangle = ||Ax||^2 \geq 0$ por otro. Entonces

$$\lambda||x||^2 = ||Ax||^2 \geq 0 \Rightarrow \lambda = \frac{||Ax||^2}{||x||^2}$$

ya que $||x||^2 \neq 0$ por ser x vector propio y por lo tanto $\lambda \geq 0$.

- Los valores propios no nulos de AA' y de $A'A$ coinciden (ver pizarrón).

Teorema Espectral para matrices simétricas:
Sea $A \in \mathcal{M}_{n\times n}$ simétrica. Entonces A es diagonalizable en una base ortonormal de $\mathbb{R}^n$ y podemos escribir:

$$A = UDU' \quad \text{con } U \in \mathcal{M}_{n\times n} \text{ ortogonal}$$

$$A = (u_1 \ u_2 \ \dots \ u_n) \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_n \end{pmatrix} \begin{pmatrix} u_1' \\ u_2' \\ \vdots \\ u_n' \end{pmatrix}$$

$$= (u_1 \ u_2 \ \dots \ u_n) \begin{pmatrix} \lambda_1 u_1' \\ \lambda_2 u_2' \\ \vdots \\ \lambda_n u_n' \end{pmatrix} = \sum_{i=1}^{n} \lambda_i u_i u_i'$$

La descomposición matricial $A = \sum_{i=1}^{n} \lambda_i u_i u_i'$ en suma de n matrices se llama *descomposición espectral* de A.

Importante: no confundir con $\sum_{i=1}^{n} \lambda_i u_i' u_i = \sum_{i=1}^{n} \lambda_i = \lambda_1 + \dots + \lambda_n \in \mathbb{R}$

Descomposición en valores singulares

Si $A \in \mathcal{M}_{n\times p}$ con $n \geq p$ tiene rango r entonces A se puede descomponer como

$$A = UDV'$$

con

- $V \in \mathcal{M}_{p\times r}$ cuyas columnas son los vectores propios de norma 1 **correspondientes a los valores propios no nulos de** $A'A$.
- $D \in \mathcal{M}_{r\times r} = diag(\sigma_1, \ldots, \sigma_r)$, donde $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ donde los r valores singulares de A corresponden a la raíz cuadrada de los valores propios no nulos de $A'A$ ($\sigma_j = \sqrt{\lambda_j}\ \forall j = 1, \ldots, r$), en su diagonal principal ordenados de mayor a menor.
- $U \in \mathcal{M}_{n\times r}$ cuyas columnas son los vectores propios de norma 1 **correspondientes a los valores propios no nulos de** AA'.

Entonces la *descomposición en valores singulares* de A es

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i'$$

Si $A \in \mathcal{M}_{n\times p}$ con $n \geq p$ tiene rango r entonces A se puede descomponer como

$$A = UDV'$$

con

- $V \in \mathcal{M}_{p\times r}$ cuyas columnas son los vectores propios de norma 1 **correspondientes a los valores propios no nulos de** $A'A$.
- $D \in \mathcal{M}_{r\times r} = diag(\sigma_1, \dots, \sigma_r)$, donde $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ donde los r valores singulares de A corresponden a la raíz cuadrada de los valores propios no nulos de $A'A$ ($\sigma_j = \sqrt{\lambda_j}\ \forall j = 1, \dots, r$), en su diagonal principal ordenados de mayor a menor.
- $U \in \mathcal{M}_{n\times r}$ cuyas columnas son los vectores propios de norma 1 **correspondientes a los valores propios no nulos de** AA'.

Entonces la *descomposición en valores singulares* de A es

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^{'}$$

Obs: todo se basa en que $A'A$ y AA' tienen los mismos valores propios no nulos y que si v es vector propio de $A'A$ entonces Av es vector propio de AA' asociado al mismo valor propio. En efecto, por construcción, si λ_j es valor propio de $A'A$ asociado a w_j entonces λ_j es valor propio de AA' asociado a u_j ya que

$$AA'(u_j) = AA'\left(\frac{Av_j}{\sigma_j}\right) = A\left(\frac{\lambda_j v_j}{\sigma_j}\right) = \lambda_j u_j$$

Descomposición SVD - Ejemplo

Ejemplo: Hallar la descomposición SVD de $A = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix}$

1. La matriz $A'A$ es $\begin{pmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{pmatrix}$ con valores y subespacios propios

$$\left\{ \begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix} \right\} \leftrightarrow 0, \quad \left\{ \begin{pmatrix} 1 \\ -1 \\ 4 \end{pmatrix} \right\} \leftrightarrow 9, \quad \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right\} \leftrightarrow 25$$

2.
   - $D = \begin{pmatrix} 5 & 0 \\ 0 & 3 \end{pmatrix}$ $\qquad V = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{18} \\ -1/\sqrt{2} & -1/\sqrt{18} \\ 0 & 4/\sqrt{18} \end{pmatrix}$
   - Haciendo $\begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix} \times \frac{1}{5} = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$,

     $\begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix} \begin{pmatrix} 1/\sqrt{18} \\ -1/\sqrt{18} \\ 4/\sqrt{18} \end{pmatrix} \times \frac{1}{3} = \begin{pmatrix} \frac{1}{2}\sqrt{2} \\ -\frac{1}{2}\sqrt{2} \end{pmatrix}$.

     La matriz U es $U = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$
   - Por lo tanto

     $A = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{18} & -1/\sqrt{18} & 4/\sqrt{18} \end{pmatrix} = UDV'$

Si $n \geq p$, entonces podemos considerar $A' \in \mathcal{M}_{p \times n}$ y $n \geq p$ (como en el caso que demostramos) y $A' = UDV'$. Por lo tanto $A = VDU'$ (¡y la U' es la V' y la V es la U!). Por ejemplo, sea $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$. Acá $n \geq p$, entonces $A'A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ tiene como valores propios 3 y 1 asociados respectivamente a $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ y $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$. Entonces

$$A = \begin{pmatrix} 2/3\sqrt{3} & 0 \\ 1/3\sqrt{3} & -1/\sqrt{2} \\ 1/3\sqrt{3} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

Plan

Función escalar

1. Si $\mathbf{x} = (x_1, x_2, x_3)$ y $f(\mathbf{x}) = f(x_1, x_2, x_3) = 4x_1 - 3x_2 + 2x_3$ entonces

$$\frac{\partial f}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{pmatrix} \begin{pmatrix} 4 \\ -3 \\ 2 \end{pmatrix}$$

Función escalar

1. Si $\mathbf{x} = (x_1, x_2, x_3)$ y $f(\mathbf{x}) = f(x_1, x_2, x_3) = 4x_1 - 3x_2 + 2x_3$ entonces

$$\frac{\partial f}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{pmatrix} \begin{pmatrix} 4 \\ -3 \\ 2 \end{pmatrix}$$

2. Si $f(\mathbf{x}) = a'\mathbf{x}$ entonces $\frac{\partial f}{\partial \mathbf{x}} = a$

Función escalar

1. Si $\mathbf{x} = (x_1, x_2, x_3)$ y $f(\mathbf{x}) = f(x_1, x_2, x_3) = 4x_1 - 3x_2 + 2x_3$ entonces
$$\frac{\partial f}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{pmatrix} \begin{pmatrix} 4 \\ -3 \\ 2 \end{pmatrix}$$

2. Si $f(\mathbf{x}) = a'\mathbf{x}$ entonces $\frac{\partial f}{\partial \mathbf{x}} = a$

3. Si A es simétrica y $f(\mathbf{x}) = \mathbf{x}'A\mathbf{x}$ entonces $\frac{\partial f}{\partial \mathbf{x}} = 2A\mathbf{x} = \begin{pmatrix} 2a_1^{'}\mathbf{x} \\ 2a_2^{'}\mathbf{x} \\ \vdots \\ 2a_n^{'}\mathbf{x} \end{pmatrix}$

Esto se debe a que $\mathbf{x}'A\mathbf{x} = \sum_{i=1}^{n} a_{ii}x_i^2 + 2\sum_{j>i} a_{ij}x_ix_j$ y $\frac{\partial f}{\partial x_1} = 2a_{11}x_1 + 2a_{12}x_2 + 2a_{1n}x_n = 2a_1^{'}\mathbf{x}$

Función vectorial

1. Si $\mathbf{y} = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{pmatrix}$ entonces $\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial f_1(\mathbf{x})}{\partial \mathbf{x}} & \frac{\partial f_2(\mathbf{x})}{\partial \mathbf{x}} & \dots & \frac{\partial f_n(\mathbf{x})}{\partial \mathbf{x}} \end{pmatrix}$ o sea

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \dots & \frac{\partial f_n(\mathbf{x})}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial f_1(\mathbf{x})}{\partial x_n} & \dots & \frac{\partial f_n(\mathbf{x})}{\partial x_n} \end{pmatrix}$$

Función vectorial

1. Si $\mathbf{y} = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{pmatrix}$ entonces $\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial f_1(\mathbf{x})}{\partial \mathbf{x}} & \frac{\partial f_2(\mathbf{x})}{\partial \mathbf{x}} & \dots & \frac{\partial f_n(\mathbf{x})}{\partial \mathbf{x}} \end{pmatrix}$ o sea

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \dots & \frac{\partial f_n(\mathbf{x})}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial f_1(\mathbf{x})}{\partial x_n} & \dots & \frac{\partial f_n(\mathbf{x})}{\partial x_n} \end{pmatrix}$$

2. Si $\mathbf{y} = A\mathbf{x}$ entonces $\frac{\partial A\mathbf{x}}{\partial \mathbf{x}} = A'$

En efecto si $A = \begin{pmatrix} a_1^{'} \\ \vdots \\ a_n^{'} \end{pmatrix}$ entonces $\mathbf{y} = A\mathbf{x} = \begin{pmatrix} a_1^{'}\mathbf{x} \\ \vdots \\ a_n^{'}\mathbf{x} \end{pmatrix}$ y como $\frac{\partial f_i(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial a_i^{'}\mathbf{x}}{\partial \mathbf{x}} = a_i$

entonces $\frac{\partial A\mathbf{x}}{\partial \mathbf{x}} = (a_1 \ a_2 \ \cdots \ a_n) = A'$

Para cada una de las variables (columnas) $j = 1, \ldots, p$ tenemos estadísticos de descripción:

- Media de la variable j: $\overline{x_j} = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \in \mathbb{R}$
- Varianza de la variable j: $s_j^2 = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x_j})^2$
- Desviación estándar de la variable j: s_j

El vector de medias de X es el vector $\bar{\mathbf{x}}$ en el cual cada coordenada de $\bar{\mathbf{x}}$ es la media de la columna j, es decir:

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x_i} = \begin{pmatrix} \overline{x_1} \\ \overline{x_2} \\ \vdots \\ \overline{x_p} \end{pmatrix} = \frac{1}{n}\mathbf{X}'\mathbf{1} \in \mathbb{R}^p$$

donde $\mathbf{1}$ es el vector de $\mathbb{R}^n$ con todas sus entradas iguales a 1.
Observar que:

- $\sum_{i=1}^{n}(\mathbf{x_i} - \bar{\mathbf{x}}) = 0_{\mathbb{R}^p}$.
- si $\overline{x_j} = 0$ entonces $s_j = \sqrt{\frac{1}{n}}||x_j||$

Plan

Matriz de varianzas y covarianzas

La relación **lineal** entre dos variables se mide por la covarianza.

La covarianza entre las variables x_j y x_k se define como:

$$Cov(x_j, x_k) = s_{jk} = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x_j})(x_{ik} - \overline{x_k}) = \frac{1}{n}\left\langle \begin{pmatrix} x_{1j} - \overline{x_j} \\ x_{2j} - \overline{x_j} \\ \vdots \\ x_{nj} - \overline{x_j} \end{pmatrix}, \begin{pmatrix} x_{1k} - \overline{x_k} \\ x_{2k} - \overline{x_k} \\ \vdots \\ x_{nk} - \overline{x_k} \end{pmatrix} \right\rangle$$

- Si suponemos que $\overline{x_j} = \overline{x_k} = 0$ y que $s_j = s_k = 1$ entonces

  $$x_j' x_k = \sum_{i=1}^{n}(x_{ij} - \overline{x_j})(x_{ik} - \overline{x_k}) = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x_j})(x_{ik} - \overline{x_k})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x_j})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ik} - \overline{x_k})^2}} = r_{jk}$$

  donde r_{jk} es el coeficiente de correlación de las variables j y k.

- Si suponemos que $\overline{x_j} = \overline{x_k} = 0$ se tiene que:

  $$cos(x_j, x_k) = \frac{x_j' x_k}{||x_j'||||x_k||} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x_j})(x_{ik} - \overline{x_k})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x_j})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ik} - \overline{x_k})^2}} = r_{jk}$$

Consideramos distintas matrices:

- **Matriz de datos:**

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_n' \end{pmatrix} \in \mathcal{M}_{n\times p}$$

- **Matriz de varianzas y covarianzas:**

$$S = ((s_{jk})) = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x_i} - \bar{\mathbf{x}})(\mathbf{x_i} - \bar{\mathbf{x}})' = \begin{pmatrix} s_1^2 & s_{12} & \dots & s_{1p} \\ s_{21} & s_2^2 & \dots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \dots & s_p^2 \end{pmatrix} \in \mathcal{M}_{p\times p}$$

  Explicitar!

- **Matriz de correlaciones:**

$$R = \begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{pmatrix} \in \mathcal{M}_{p\times p}$$

siendo $r_{ij} = \frac{s_{ij}}{s_i s_j}$.

Matriz de varianzas y covarianzas en R

Un ejemplo en R. Datos de 150 flores de lirio.

- Adjunto los datos: attach(iris)
- Considero la parte de la base con las variables independientes x=iris[,1:4]
- Calculo la media de cada variable: colMeans(x)
- Calculo la varianza o el desvío de cada variable: apply(x,2,var) ; apply(x,2,sd)
- La matriz de varianzas y covarianzas es: S=cov(x) o S=var(x)
- La matriz de correlaciones es: R=cor(x)
- Buscar los valores y vectores propios de S: eigen(S)

Matriz de datos centrados

Muchas veces consideraremos la matriz de datos centrados. La misma se obtiene de la matriz de datos original restando en cada columna la media de la variable correspondiente. Formalmente, si $\mathbf{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathcal{M}_{n\times 1}$, la matriz de datos centrados es:

$$\tilde{\mathbf{X}} = \mathbf{X} - \mathbf{1}\overline{\mathbf{x}}' = \mathbf{X} - \frac{1}{n}\mathbf{1}(\mathbf{1}'\mathbf{X}) = \underbrace{\left(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)}_{=P} X = P\mathbf{X}$$

Ejercicio: Probar que P es simétrica, idempotente ($PP = P$) y de rango $= n-1$ (pues es ortogonal al subespacio generado por el vector $\mathbf{1}$).

Ejercicio: Probar que

$$S = \frac{1}{n}\tilde{\mathbf{X}}'\tilde{\mathbf{X}} = \frac{1}{n}\mathbf{X}'P\mathbf{X}$$

Observación: Se puede definir la matriz de varianzas corregida $\widehat{S} = \frac{1}{n-1}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$ ($n-1$ en lugar de n) a los efectos de tener un estimador insesgado de la matriz de la poblacin. Este cambio en la divisin tiene el mismo origen que en el caso univariado: hay $n-1$, y no n, desviaciones estandares independientes ya que los vectores de desviación están ligados por

$$\sum_{i=1}^{n}(\mathbf{x_i} - \overline{\mathbf{x}}) = 0$$

Veamos una serie de propiedades de la matriz de varianzas y covarianzas

$$S = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x_i} - \overline{\mathbf{x}})(\mathbf{x_i} - \overline{\mathbf{x}})' = \begin{pmatrix} s_1^2 & s_{12} & \dots & s_{1p} \\ s_{21} & s_{22} & \dots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \dots & s_p^2 \end{pmatrix} \in \mathcal{M}_{p\times p}$$

1. S es simétrica y tiene el mismo rango que $\tilde{\mathbf{X}}$.
2. S es diagonalizable en una base ortonormal de $\mathbb{R}^p$.
3. S es semidefinida positiva, o sea para todo $w \in \mathbb{R}^p$ $w'Sw \geq 0$.
4. S tiene todos sus valores propios ≥ 0.
5. $\det(S) \geq 0$, $tr(S) \geq 0$.

Estas propiedades son MUY importantes y las utilizaremos frecuentemente en los captulos siguientes.

S es simétrica y tiene el mismo rango que $\tilde{\mathbf{X}}$

- S es simétrica pues $S = \frac{1}{n}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$ y la matriz $\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$ es simétrica. Acá estamos usando que AA' y $A'A$ son simétricas.
- $rg(S) = rg(\tilde{X})$ pues $S = \frac{1}{n}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$ y acá usamos que $rg(A'A) = rg(AA') = rg(A)$

Propiedades de la matriz de varianza y covarianzas

S es simétrica y tiene el mismo rango que $\tilde{\mathbf{X}}$

- S es simétrica pues $S = \frac{1}{n}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$ y la matriz $\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$ es simétrica. Acá estamos usando que AA' y $A'A$ son simétricas.
- $rg(S) = rg(\tilde{X})$ pues $S = \frac{1}{n}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$ y acá usamos que $rg(A'A) = rg(AA') = rg(A)$

S es diagonalizable en una base ortonormal de $\mathbb{R}^p$

Al ser S simétrica, existe una base ortonormal de $\mathbb{R}^p$ formada por vectores propios de S.

S es semidefinida positiva, o sea para todo $w \in \mathbb{R}^p$ $w'Sw \geq 0$

Esta propiedad es la análoga al caso univariante para el cual la varianza es un número no negativo. Debemos probar que si $w \in \mathbb{R}^p$ es un vector cualquiera entonces $w'Sw \geq 0$. Sea $w \in \mathbb{R}^p$ *cualquiera*. Para cada $i = 1, \dots, n$ defino

$$v_i = w'(\mathbf{x_i} - \overline{\mathbf{x}}) \in \mathbb{R}$$

Entonces la media de esta variable es $\overline{v} = \frac{1}{n}\sum_{i=1}^{n} v_i = \frac{1}{n}\sum_{i=1}^{n} w'(\mathbf{x_i} - \overline{\mathbf{x}}) = \frac{1}{n} w' \sum_{i=1}^{n}(\mathbf{x_i} - \overline{\mathbf{x}}) = 0$

Por otro lado, su varianza es

$$Var(v) = \frac{1}{n}\sum_{i=1}^{n} v_i^2 = \underbrace{\frac{1}{n}\sum_{i=1}^{n} w'(\mathbf{x_i} - \overline{\mathbf{x}})(\mathbf{x_i} - \overline{\mathbf{x}})'w}_{w'Sw} \geq 0 \Rightarrow S \text{ es semidefinida positiva.}$$

La matriz S tiene todos sus valores propios no negativos

En efecto si λ es un valor propio de S (tiene pues S es simétrica) y $\mathbf{x}$ es un vector propio asociado (por lo tanto no nulo) entonces

$$0 \leq \mathbf{x}' S\mathbf{x} = \mathbf{x}'(\lambda \mathbf{x}) = \lambda \mathbf{x}'\mathbf{x} = \lambda \underbrace{||x||^2}_{>0} \Rightarrow \lambda \geq 0$$

La traza de S y el determinante de S son no negativos.

En efecto como los valores propios de S son todos no negativos, al ser la traza de S la suma de los valores propios de S y el determinante de S el producto de los valores propios se deduce el resultado.

Plan

Eliminación de variables redundantes

Veamos como es posible reducir la cantidad de variables de la matriz de datos cuando una variable es combinación lineal de otras.

Lema: $\{w \in \mathbb{R}^p : w'Sw = 0\} = N(S)$.

Haremos la prueba en clase. En efecto:

- es claro que $N(S) \subset \{w : w'Sw = 0\}$.
- Sea $w \in \{w \in \mathbb{R}^p : w'Sw = 0\}$, entonces descomponiendo w en una base ortonormal de $\mathbb{R}^p$ en la que S es diagonalizable se prueba la otra inclusión.

Proposición: $N(S) = N(\tilde{\mathbf{X}})$

Dem:

$$w \in N(S) \Leftrightarrow Sw = 0_{\mathbb{R}^p} \Leftrightarrow w'\tilde{\mathbf{X}}'\tilde{\mathbf{X}}w = 0$$
$$\Leftrightarrow (\tilde{\mathbf{X}}\mathbf{w})'\tilde{\mathbf{X}}w = 0 \Leftrightarrow ||\tilde{\mathbf{X}}w||^2 = 0 \Leftrightarrow \tilde{\mathbf{X}}w = 0_{\mathbb{R}^n} \Leftrightarrow w \in N(\tilde{\mathbf{X}})$$

Conclusión: si un vector w se encuentra en el núcleo de S, sus coordenadas son los coeficientes de la combinación lineal existente entre las variables de la matriz de datos.

Eliminación de variables redundantes: ejemplo

Consideremos la siguiente matriz de varianzas y covarianzas (ejemplo 3.3 del libro de Peña):

$$S = \begin{pmatrix} 0{,}0947 & 0{,}0242 & 0{,}0054 & 0{,}0594 \\ 0{,}0242 & 0{,}0740 & 0{,}0285 & 0{,}0491 \\ 0{,}0054 & 0{,}0285 & 0{,}0838 & 0{,}0170 \\ 0{,}0594 & 0{,}0285 & 0{,}0170 & 0{,}0543 \end{pmatrix}$$

cuyos valores propios son 0.172, 0.08762, 0.0461 y 0.00005 (el rango es aprox. 3).

Un vector propio asociado al valor propio cero es $(0{,}408, 0{,}408, 0, -0{,}816)$ que se puede reescribir como $(0{,}5, 0{,}5, 0, -1)$ por lo que la 4ta variable de X es el promedio de las dos primeras.

Obs.: El razonamiento anterior se puede extender para cualquier nmero de valores propios nulos: si S tiene rango h entonces existen $p - h$ variables redundantes que podemos eliminar y los vectores asociados a estos valores propios indican la CL de estas variables redundantes (ver páginas 76 y 77 del libro de Peña).

Recordamos que el coeficiente de correlacin entre las variables k y j es

$$r_{jk} = \frac{s_{jk}}{s_j s_k}$$

La matriz de correlacin es:

$$R = \begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{pmatrix} \in \mathcal{M}_{p\times p}$$

Si $D = D(S)$ es la matriz diagonal cuya diagonal son las varianzas de las variables. Entonces R se puede escribir como

$$R = D^{-1/2} S D^{-1/2}$$

y por lo tanto $S = D^{1/2} R D^{1/2}$.

Lo anterior implica que, al igual que S, la matriz R es semidefinida positiva.

Si $S = \begin{pmatrix} s_1^2 & s_{12} \\ s_{21} & s_2^2 \end{pmatrix}$ entonces $D = \begin{pmatrix} s_1^2 & 0 \\ 0 & s_2^2 \end{pmatrix}$ y $D^{-1/2} = \begin{pmatrix} \frac{1}{s_1} & 0 \\ 0 & \frac{1}{s_2} \end{pmatrix}$ y

$$D^{-1/2} S D^{1/2} = \begin{pmatrix} \frac{1}{s_1} & 0 \\ 0 & \frac{1}{s_2} \end{pmatrix} \begin{pmatrix} s_1^2 & s_{12} \\ s_{21} & s_2^2 \end{pmatrix} \begin{pmatrix} \frac{1}{s_1} & 0 \\ 0 & \frac{1}{s_2} \end{pmatrix} = \begin{pmatrix} \frac{s_1^2}{s_1 s_1} & \frac{s_{12}}{s_1 s_2} \\ \frac{s_{12}}{s_1 s_2} & \frac{s_2^2}{s_2 s_2} \end{pmatrix} = R$$

Dependencia de una variable y el resto: la regresión múltiple

Otra manera de ver si una variable es dependiente de las demás es de medir su grado de dependencia mediante una regresión lineal. Si $x_j = y$ es la variable que queremos explicar a partir de las demás variables el predictor lineal que consideramos es

$$\widehat{y}_i = \overline{y} + \widehat{\beta}_1(x_{i1} - \overline{x}_1) + \widehat{\beta}_2(x_{i2} - \overline{x}_2) + \cdots + \widehat{\beta}_p(x_{ip} - \overline{x}_p) \quad i = 1, \ldots, n$$

Los $p-1$ coeficientes $\widehat{\beta}_k$ se toman de manera que la ecuación proporcione, en promedio, la mejor predicción posible de los valores de y_i. Éstos se calculan minimizando

$$||e||^2 = e'e = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2$$

Dependencia de una variable y el resto: la regresión múltiple

Otra manera de ver si una variable es dependiente de las demás es de medir su grado de dependencia mediante una regresión lineal. Si $x_j = y$ es la variable que queremos explicar a partir de las demás variables el predictor lineal que consideramos es

$$\widehat{y}_i = \overline{y} + \widehat{\beta}_1(x_{i1} - \overline{x}_1) + \widehat{\beta}_2(x_{i2} - \overline{x}_2) + \cdots + \widehat{\beta}_p(x_{ip} - \overline{x}_p) \quad i = 1, \ldots, n$$

Los $p-1$ coeficientes $\widehat{\beta}_k$ se toman de manera que la ecuación proporcione, en promedio, la mejor predicción posible de los valores de y_i. Éstos se calculan minimizando

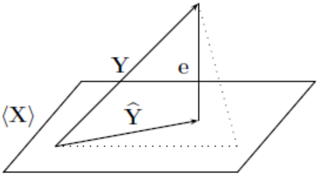
$$||e||^2 = e'e = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2$$

Interpretación geométrica: Recordar que si S es un subespacio de $\mathbb{R}^n$ y P_S la proyección ortogonal sobre S, entonces

$$||v - P_S(v)|| \leq ||v - s|| \;\; \forall\, s \in S$$

Por lo tanto es mínimo $e'e = ||\mathbf{Y} - \tilde{\mathbf{X}}\beta||^2$ cuando

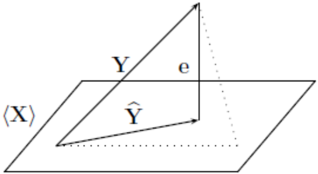
$$\tilde{\mathbf{X}}\beta = P_{\langle \tilde{\mathbf{X}} \rangle}(Y) = \widehat{\mathbf{Y}}$$



donde $\langle \tilde{\mathbf{X}} \rangle$ es el subespacio generado por las columnas de la matriz de datos centrada, β es el vector de parámetros que se quieren estimar.

De lo anterior podemos deducir que:

- $e = \mathbf{Y} - \widehat{\mathbf{Y}}$ es ortogonal a $\langle \tilde{\mathbf{X}} \rangle$
- $\tilde{\mathbf{X}}'\mathbf{e} = \mathbf{0}$ es decir $\sum_{i=1}^{n} e_i x_{ik} = 0 \;\; k = 1, \dots, p, \; k \neq j$



Entonces multiplicando por $\tilde{\mathbf{X}}'$ la igualdad $\widehat{\mathbf{Y}} = \tilde{\mathbf{X}}\beta$ siendo $\widehat{\mathbf{Y}}$ la variable que se quiere prever, se tiene

$$\tilde{\mathbf{X}}'\widehat{\mathbf{Y}} = \tilde{\mathbf{X}}'\tilde{\mathbf{X}}\beta$$

y un estimador para β es

$$\widehat{\beta} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\mathbf{Y}$$

Observación: En el despeje anterior supusimos que la matriz $\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$ es invertible. Esto es cierto si el rango es completo, es decir si las columnas de $\tilde{\mathbf{X}}$ forman un conjunto LI (para eso recordar que $rg(A'A) = rg(A)$).

Dependencia de una variable y el resto: la regresión múltiple

De la igualdad $y_i - \overline{y} = \widehat{y}_i - \overline{y} + e_i$, elevando al cuadrado y sumando se obtiene que

$$\underbrace{\sum_{i=1}^{n}(y_i - \overline{y})^2}_{\text{Variación total VT o inicial de los datos}} = \underbrace{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}_{\text{Variación no explicada VNE por la regresión}} + \underbrace{\sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2}_{\text{Variación explicada VE por la regresión}}$$

El coeficiente de correlación múltiple entre la variable j y el resto es

$$R^2_{j,1,\ldots,p} = \frac{VE}{VT} = 1 - \frac{VNE}{VT} = 1 - \frac{\frac{1}{n}\sum\limits_{i=1}^{n} e_i^2}{s_j^2}$$

Es posible probar (ver Peña pág. 89) que si notamos por $s_{jj} = s_j^2$ el elemento de la entrada jj de S, y por s^{jj} el elemento de la entrada jj de S^{-1}, entonces

$$s^{jj} = \frac{1}{\frac{1}{n}\sum\limits_{i=1}^{n} e_i^2}$$

y por lo tanto el coeficiente de correlación múltiple de la variable j con respecto de todas las otras se obtiene de S y S^{-1} mediante

$$R^2_{j,1,\ldots,p} = 1 - \frac{1}{s^{jj}s_{jj}}$$

Multicolinealidad

El *Factor de Inflación de la Varianza* de la variable x_j se define como

$$\mathrm{FIV}_j = \frac{1}{1 - R_j^2}$$

donde R_j^2 es el de la regresión de la variable x_j contra todas las demás variables.

- Si $R_j^2 = 0$, es decir cuando x_j no depende linealmente de las demás variables, entonces $\mathrm{FIV}_j = 1$.
- Si $R_j^2 \neq 0$ entonces $\mathrm{FIV}_j > 1$ y si $R_j^2 \approx 1$ entonces FIV_j es grande.

Cuando $\mathrm{FIV}_j > 10$ el problema de multicolinealidad es grande. Posibles soluciones a la multicolinealidad son:

- Eliminar la variable problématica.
- Combinar las variables colineales en una única, estandarizándolas y promediándolas

Código cálculo de los FIV's

```
vif_calc<-function(Xmat){
VIF<-numeric()
for(i in 1:ncol(Xmat)){
Xmat_Y<-Xmat[,i]
dataMAT<-cbind(Xmat_Y, Xmat[,-i])
R2<-summary(lm(Xmat_Y~.,data=dataMAT, na.action="na.exclude"))$r.squared

VIF[i]<-1/(1-R2)
}
names(VIF)<-colnames(Xmat)
print(VIF)
}

> vif_calc(iris[,1:4])
Sepal.Length  Sepal.Width Petal.Length  Petal.Width
    7.072722     2.100872    31.261498    16.090175
```

Ejemplo 3.1 *La tabla A.5 del Apéndice de Datos, MEDIFIS, presenta ocho variables físicas tomadas en un grupo de 27 estudiantes. Las variables son sexo (sex con 0 para mujer, 1 para varón), estatura (est, en cm.), peso (pes, en kgr.), longitud de pie (lpie, en cm), longitud de brazo (lbra, en cm), anchura de la espalda (aes, en cm), diámetro de cráneo (dcr, en cm) y longitud entre la rodilla y el tobillo (lrt, en cm).*

Ejemplo 3.9 *La matriz de correlación para las 7 variables físicas, tabla A.5, MEDIFIS, del ejemplo 1.1. se presenta en la tabla 1.5. Las variables aparecen en el orden del ejemplo 1.1*

$$\mathbf{R} = \begin{bmatrix} 1 & 0.83 & 0.93 & 0.91 & 0.84 & 0.59 & 0.84 \\ 0.83 & 1 & 0.85 & 0.82 & 0.84 & 0.62 & 0.72 \\ 0.93 & 0.85 & 1 & 0.85 & 0.80 & 0.55 & 0.85 \\ 0.91 & 0.82 & 0.85 & 1 & 0.80 & 0.48 & 0.76 \\ 0.84 & 0.84 & 0.80 & 0.80 & 1 & 0.63 & 0.63 \\ 0.59 & 0.62 & 0.55 & 0.48 & 0.63 & 1 & 0.56 \\ 0.84 & 0.72 & 0.85 & 0.76 & 0.63 & 0.56 & 1 \end{bmatrix}$$

Se observa que la máxima correlación aparece entre la primera y la tercera variable (estatura y longitud del pie) y es 0,93. La mínima es entre la longitud del brazo y el diámetro del cráneo (0,48). En general las correlaciones más bajas aparecen entre el diámetro del cráneo y el resto de las variables. La matriz S^{-1} *es:*

$$\begin{matrix} 0.14 & 0.01 & -0.21 & -0.11 & -0.07 & -0.05 & -0.07 \\ 0.01 & 0.04 & -0.08 & -0.03 & -0.04 & -0.04 & -0.00 \\ -0.21 & -0.08 & 1.26 & 0.06 & -0.05 & 0.18 & -0.29 \\ -0.11 & -0.03 & 0.06 & 0.29 & -0.04 & 0.13 & -0.04 \\ -0.07 & -0.04 & -0.05 & -0.04 & 0.34 & -0.13 & 0.15 \\ -0.05 & -0.04 & 0.18 & 0.13 & -0.13 & 0.64 & -0.15 \\ -0.07 & -0.00 & -0.29 & -0.04 & 0.15 & -0.15 & 0.50 \end{matrix}$$

y utilizando los elementos diagonales de esta matriz y de la matriz S podemos calcular las correlaciones múltiples al cuadrado de cada variable con el resto como sigue: (1) multiplicamos los elementos diagonales de las matrices S y S^{-1}. El resultado de esta operacion es el

vector (14.3672, 5.5415, 9.9898, 6.8536, 5.3549, 2.0784, 4.7560). (2) A continuación, calculamos las inversas de estos elementos, para obtener (0.0696 0.1805 0.1001 0.1459 0.1867 0.4811 0.2103). Finalmente, restamos a uno estos coeficientes para obtener (0.9304, 0.8195, 0.8999, 0.8541, 0.8133, 0.5189, 0.7897) y estos son los coeficientes de correlación múltiple entre cada variable y el resto. Vemos que la variable más previsible por las restantes es la estatura,($R^2 = 0.9304$) , despues el pié ($R^2 = 0.8999$) y luego la longitud del brazo ($R^2 = 0.8541$). La menos predecible es dcr, que tiene un coeficiente de correlación múltiple con el resto de 0.5189, o en otros términos, el resto de las variables explica el 52% de la variabilidad de esta variable.

La ecuación para prever la estatura en función del resto de las variables se obtiene facilmente con cualquier programa de regresión. El resultado es

$$est = 0.9 - 0.094\ peso + 1.43 pie + 0.733\ lbr + 0.494\ aes + 0.347\ dcr + 0.506\ lrt$$

que es la ecuación que permite prever con menor error la estatura de una persona dadas el resto de las medidas. El R^2 de esta regresión es = 0,93, resultado que habíamos obtenido anteriormente. La ecuación para prever la longitud del pié es:

$$pie = 8.14 + 0.162\ est + 0.0617\ pes - 0.051\ lbr + 0.037\ aes - 0.144\ dcr + 0.229 lrt$$

Plan

Medidas globales de variabilidad

Es un resumen de la matriz de varianzas y covarianzas.

- Varianza total

$$T = tr(S) = \sum_{i=1}^{p} s_i^2$$

- Varianza media

$$\overline{s}^2 = \frac{1}{p}\sum_{i=1}^{p} s_i^2$$

Estas dos medidas no toman la estructura de dependencia entre las variables

Medidas globales de variabilidad

Es un resumen de la matriz de varianzas y covarianzas.

- Varianza total

$$T = tr(S) = \sum_{i=1}^{p} s_i^2$$

- Varianza media

$$\bar{s}^2 = \frac{1}{p}\sum_{i=1}^{p} s_i^2$$

Estas dos medidas no toman la estructura de dependencia entre las variables

- Varianza generalizada

$$VG = |S|$$

Siempre es mayor o igual a 0 ya que $|S| \geq 0$ y es una medida del volumen ocupado por el conjunto de datos (grado de dispersin). En efecto si $p = 2$ entonces como $r = \frac{s_{xy}}{s_x s_y}$ se tiene que:

$$S = \begin{pmatrix} s_x^2 & r s_x s_y \\ r s_x s_y & s_y^2 \end{pmatrix}$$

y la desviacin tipíca generalizada es $\sqrt{|S|} = s_x s_y \sqrt{1 - r^2}$. Si las variables son **independientes** la mayora de sus valores van a estar dentro de un rectángulo de lados $6s_x$ y $6s_y$ ya que por el teorema de Tchebyshev entre la media y 3 desviaciones tipícas deben estar entre aprox. el 90 y el 100 % de los datos ($P(|X - \mu_x| > 3s_x) \leq \frac{1}{9}$) Por lo que el área ocupado por ambas variables es proporcional a $s_x s_y$. A medida que la relación entre ambas variables es fuerte, el rea ocupado va a tender a 0 alrededor de la recta de regresin.

No sirve para comparar conjuntos de datos con distinto número de variables.

Medidas globales de variabilidad

Es un resumen de la matriz de varianzas y covarianzas.

- Varianza total

$$T = tr(S) = \sum_{i=1}^{p} s_i^2$$

- Varianza media

$$\bar{s}^2 = \frac{1}{p}\sum_{i=1}^{p} s_i^2$$

Estas dos medidas no toman la estructura de dependencia entre las variables

- Varianza generalizada

$$VG = |S|$$

Siempre es mayor o igual a 0 ya que $|S| \geq 0$ y es una medida del volumen ocupado por el conjunto de datos (grado de dispersin). En efecto si $p = 2$ entonces como $r = \frac{s_{xy}}{s_x s_y}$ se tiene que:

$$S = \begin{pmatrix} s_x^2 & rs_x s_y \\ rs_x s_y & s_y^2 \end{pmatrix}$$

y la desviacin tipíca generalizada es $\sqrt{|S|} = s_x s_y \sqrt{1 - r^2}$. Si las variables son **independientes** la mayora de sus valores van a estar dentro de un rectángulo de lados $6s_x$ y $6s_y$ ya que por el teorema de Tchebyshev entre la media y 3 desviaciones tipícas deben estar entre aprox. el 90 y el 100 % de los datos ($P(|X - \mu_X| > 3s_x) \leq \frac{1}{9}$) Por lo que el área ocupado por ambas variables es proporcional a $s_x s_y$. A medida que la relación entre ambas variables es fuerte, el rea ocupado va a tender a 0 alrededor de la recta de regresin.

No sirve para comparar conjuntos de datos con distinto número de variables.

- Varianza efectiva

$$VE = |S|^{1/p}$$

Distancia entre individuos

- La distancia de Euclides depende mucho de las unidades de medida de las variables. En efecto si queremos comparar la distancia entre las personas $A(1{,}80; 80)$, $B(1{,}70 : 72)$ y $C(1{,}65; 81)$ y

$$d^2(A,B) = (1{,}80 - 1{,}70)^2 + (80-72)^2 = 64{,}01 \quad \text{y}\, d^2(A,C) = 1{,}225$$

Por lo que A est más cerca de C. Si considero las mismas personas pero expreso en cm su altura

$$d^2(A,B) = 164 \quad \text{y}\, d^2(A,C) = 226$$

y obtenemos que A está más cerca de B esta vez...

- La distancia de Euclides depende mucho de las unidades de medida de las variables. En efecto si queremos comparar la distancia entre las personas $A(1{,}80; 80)$, $B(1{,}70 : 72)$ y $C(1{,}65; 81)$ y

$$d^2(A,B) = (1{,}80 - 1{,}70)^2 + (80-72)^2 = 64{,}01 \quad \text{y}\, d^2(A,C) = 1{,}225$$

Por lo que A est más cerca de C. Si considero las mismas personas pero expreso en cm su altura

$$d^2(A,B) = 164 \quad \text{y}\, d^2(A,C) = 226$$

y obtenemos que A está más cerca de B esta vez...

- Una manera de remediar a eso es de dividir cada variables por un termino que elimine el efecto de escala y considerar este tipo de métrica

$$d_{ij} = \left[(\mathbf{x_i} - \mathbf{x_j})' M^{-1} (\mathbf{x_i} - \mathbf{x_j})\right]^{1/2}$$

donde M es una matriz no singular, simétrica y definida positiva (¡lo que le pedimos a la matriz de varianzas-covarianzas de rango p!).

Distancia de Mahalanobis

La distancia de Mahalanobis entre dos puntos es

$$d_{ij} = \left[(\mathbf{x_i} - \mathbf{x_j})' S^{-1} (\mathbf{x_i} - \mathbf{x_j})\right]^{1/2}$$

siendo S la matriz de varianzas y covarianzas.

Distancia de Mahalanobis

La distancia de Mahalanobis entre dos puntos es

$$d_{ij} = \left[(\mathbf{x_i} - \mathbf{x_j})' S^{-1} (\mathbf{x_i} - \mathbf{x_j})\right]^{1/2}$$

siendo S la matriz de varianzas y covarianzas.
La distancia de Mahalanobis entre un punto y su vector de medias es

$$d_i = \left[(\mathbf{x_i} - \overline{\mathbf{x}})' S^{-1} (\mathbf{x_i} - \overline{\mathbf{x}})\right]^{1/2}$$

siendo S la matriz de varianzas y covarianzas.

La distancia de Mahalanobis entre dos puntos es

$$d_{ij} = \left[(\mathbf{x_i} - \mathbf{x_j})' S^{-1} (\mathbf{x_i} - \mathbf{x_j}) \right]^{1/2}$$

siendo S la matriz de varianzas y covarianzas.
La distancia de Mahalanobis entre un punto y su vector de medias es

$$d_i = \left[(\mathbf{x_i} - \bar{\mathbf{x}})' S^{-1} (\mathbf{x_i} - \bar{\mathbf{x}}) \right]^{1/2}$$

siendo S la matriz de varianzas y covarianzas.

Esta distancia toma en cuenta las correlaciones entre las variables.

En el caso $p = 2$, al escribir $r = r_{12}$ y $s_{12} = rs_1 s_2$ entonces

$$S^{-1} = \frac{1}{1 - r^2} \begin{pmatrix} s_1^{-2} & -rs_1^{-1} s_2^{-1} \\ -rs_1^{-1} s_2^{-1} & s_2^{-2} \end{pmatrix}$$

y la distancia de Mahalanobis entre dos puntos (x_1, y_1) e (x_2, y_2) es

$$d_M^2 = \frac{1}{1 - r^2} \left[\frac{(x_1 - x_2)^2}{s_1^2} + \frac{(y_1 - y_2)^2}{s_2^2} - 2r \frac{(x_1 - x_2)(y_1 - y_2)}{s_1 s_2} \right]$$

Observar que si $r = 0$ es la distancia euclidea estandarizando las variables.

Por ejemplo volviendo al ejemplo anterior si $A(180, 80), B(170, 72)$ y $C(165, 81)$ si las desviaciones tipicas son $s_1 = 10$ y $s_2 = 10$ y el coeficiente de correlacin 0,7 entonces

$$d^2(A, B) = \frac{1}{0{,}51}(1+0{,}8^2-1{,}4\times0{,}8) = 1{,}02 \quad d^2(A, C) = \frac{1}{0{,}51}(1{,}5^2+0{,}1^2-1{,}4\times1{,}5\times0{,}1) = 4{,}84$$

y por lo tanto A est más cerca de B. Esta distancia toma en cuenta que como hay correlación entre la altura y el peso, si su peso disminuye proporcionalmente, el aspecto de ambos es similar (cambia el tamao global pero no la forma). Observar que C es ms bajo que A pero pesa ms, lo que implica que su fsico es bien distinto al de A.

Referencias

- D. Peña, *Analisis de Datos Multivariantes*, Mac Graw Hill, 2002.
- A. I. Izenman, *Modern Multivariate Statistical Techniques*, Springer, 2008.
- G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*, Springer, 2013.