

**Grado en Ingeniería Forestal
y del Medio Natural.
Grado en Ingeniería Geomática
y Topografía.**

**Aplicaciones del Álgebra Lineal:
prácticas de Matlab.**

Diego Delgado Bueno
Departamento de Matemáticas
Universidad de León



Resumen

El presente documento presenta una colección de problemas diseñados para mostrar algunas de las aplicaciones del Álgebra Lineal en el contexto de la asignatura de mismo nombre en los grados de Ingeniería Forestal y del Medio Natural e Ingeniería Geomática y Topografía que se imparten en el Campus de Ponferrada de la Universidad de León.

Abstract

This document presents a collection of problems designed to show some of the applications of Linear Algebra in the context of the subject of the same name in the degrees of Forestry and Natural Environment Engineering and Geomatics Engineering and Surveying taught at the Ponferrada Campus of the University of León.

Palabras clave

Álgebra lineal, Cadenas de Markov, Análisis de Componentes Principales, Valores Propios, Vectores Propios, Matlab

Key words

Lineal Algebra, Markov Chains, Principal Component Analysis, Eigenvalues, Eigenvectors, Matlab.



Este obra está bajo una [licencia de Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Índice general

I	Introducción a Matlab	5
0.1	Introducción a Matlab	7
0.2	Vectores y matrices:	7
0.3	Operaciones con vectores y matrices:	7
0.3.1	Operaciones con vectores.	8
0.3.2	Suma y resta de matrices.	8
0.3.3	Producto de matrices.	9
0.3.4	Traspuesta, determinante e inversa.	9
0.3.5	Diagonalización	10
0.4	Otros comandos de interés:	10
0.5	Ejercicios.	11
II	Problemas resueltos	13
1	Los visitantes aleatorios:	15
1.1	Resolución mediante Diagonalización	17
1.2	Resolución numérica.	18
1.3	Resolución numérica usando diagonalización.	19
2	Herencia Mendeliana.	21
2.1	Resolución mediante Diagonalización	23
2.2	Resolución numérica.	24
2.2.1	Resolución numérica usando diagonalización.	25
3	Muestreo de aves cantoras.	27
3.1	Conceptos de estadística:	27
3.2	Análisis de componentes principales:	28
3.3	Caso práctico:	29

4	Estadística de hojas de árbol	33
4.1	Conceptos de estadística:	34
4.2	Análisis de componentes principales:	35
4.3	Caso práctico.	36
III	Enunciados para los alumnos	39
0.4	Instrucciones generales.	41
1	Los visitantes aleatorios:	43
1.1	Preguntas.	44
2	Herencia Mendeliana.	47
2.1	Preguntas.	48
3	Muestreo de aves cantoras.	49
3.1	Conceptos de estadística:	50
3.2	Cuestiones.	51
4	Estadística de hojas de árbol	53
4.1	Conceptos de estadística:	54
4.2	Cuestiones.	55
5	Propagación epidémica.	57
6	Estadística de notas.	59

Parte I

Introducción a Matlab

0.1. Introducción a Matlab

La primera versión de MATLAB data de los años 70, y fue diseñada como herramienta de apoyo para los cursos de Teoría de Matrices, Álgebra Lineal y Análisis Numérico. El nombre MATLAB es un acrónimo: “Matrix Laboratory”. Hoy en día, MATLAB es un programa muy potente, con un entorno agradable, que incluye herramientas de cálculo científico y técnico y de visualización gráfica, así como un lenguaje de programación de alto nivel.

Vamos a introducir la notación de comandos básica de este lenguaje, la necesaria para poder resolver los ejercicios que se proponen en lo que sigue.

0.2. Vectores y matrices:

Veamos como se trabaja con vectores y matrices en Matlab, por ejemplo los siguientes:

- Vectores

$$\vec{v} = (1, 2, 3) \quad \vec{w} = (-1, 0, 1)$$

- Matrices

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \quad B = \begin{pmatrix} -1 & 0 & 1 \\ 3 & 1 & 0 \\ 1 & -2 & 1 \end{pmatrix}$$

Se introducen en Matlab con los comandos siguientes:

- `>> v = [1,2,3], w = [-1,0,1]`
- `>> A = [1,2,3;4,5,6;7,8,9]`
- `>> B = [-1,0,1;3,1,0;1,-2,1]`

Las variables v , w , A , y B van quedando guardadas en el espacio de trabajo con estos mismos nombres.

0.3. Operaciones con vectores y matrices:

Veamos la sintaxis de las operaciones básicas con matrices. Trabajaremos con las variables creadas anteriormente.

0.3.1. Operaciones con vectores.

La suma y resta de vectores sería:

$$\vec{v} + \vec{w} = (1, 2, 3) + (-1, 0, 1) = (0, 2, 4)$$

Y en Matlab:

$$>> v + w$$

El producto escalar estándar:

$$\vec{v} \cdot \vec{w} = (1 \ 2 \ 3) \cdot \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = 2$$

En Matlab,

$$>> v * w' = 2$$

Finalmente el producto vectorial:

$$\vec{v} \times \vec{w} = \begin{pmatrix} i & j & k \\ 1 & 2 & 3 \\ -1 & 0 & 1 \end{pmatrix} = \dots = (2, -4, 2)$$

En Matlab

$$>> \text{cross}(v, w) = (2, -4, 2)$$

0.3.2. Suma y resta de matrices.

Algebraicamente sería lo siguiente:

$$A + B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} + \begin{pmatrix} -1 & 0 & 1 \\ 3 & 1 & 0 \\ 1 & -2 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 2 & 4 \\ 7 & 6 & 6 \\ 8 & 6 & 10 \end{pmatrix}$$

Y en Matlab se operaría de forma muy similar, la siguiente:

$$>> A + B$$

O si queremos que guarde el resultado en una nueva matriz,

$$>> C = A + B$$

0.3.3. Producto de matrices.

Utilizando las matrices anteriores, el resultado del producto sería:

$$A * B = \begin{pmatrix} 8 & 4 & -4 \\ 17 & -7 & 10 \\ 26 & -10 & 16 \end{pmatrix} \gg B * A = \begin{pmatrix} 6 & 6 & 6 \\ 7 & 11 & 15 \\ 0 & 0 & 0 \end{pmatrix}$$

También se podrían multiplicar matrices fila o columna respetando la regla de las dimensiones es decir,

$$\vec{v} \cdot \vec{w}' = (1 \ 2 \ 3) \cdot \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = 2$$

Que es similar al producto escalar estándar, o bien:

$$\vec{v}' \cdot \vec{w} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \cdot (-1 \ 0 \ 1) = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -3 & 0 & 3 \end{pmatrix}$$

Debemos recordar que en general no es conmutativo. En matlab tendríamos:

$$\gg D = A * B \quad \gg D1 = B * A$$

Para el caso de las matrices cuadradas, el resultado quedaría guardado en las variables D y $D1$. Finalmente

$$\gg v * w' = 2 \quad v' * w = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -3 & 0 & 3 \end{pmatrix}$$

Para los dos casos presentados de vectores fila o columna.

0.3.4. Traspuesta, determinante e inversa.

Para poder resolver los problemas que siguen necesitamos calcular traspuestas, determinantes e inversas. La matriz traspuesta de A será la siguiente

$$A^t = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}^t = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}$$

El determinante de A y B sería el siguiente:

$$|A| = \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = 0 \quad |B| = \begin{vmatrix} -1 & 0 & 1 \\ 3 & 1 & 0 \\ 1 & -2 & 1 \end{vmatrix} = -8$$

Finalmente la inversa de B sería:

$$B^{-1} = \frac{Adj(B)^t}{|B|} = \frac{1}{4} \begin{pmatrix} -1/2 & 1 & 1/2 \\ 3/2 & 1 & -3/2 \\ 7/2 & 1 & 1/2 \end{pmatrix}$$

En Matlab se procede como sigue, traspuesta,

```
>> A'
```

determinante,

```
>> det(A) = -9,5162e - 16 >> det(B) = -8
```

inversa (hay dos maneras),

```
>> inv(B), B(-1)
```

0.3.5. Diagonalización

Matlab diagonaliza la matriz de una aplicación lineal y nos devuelve los autovalores y autovectores mediante el comando *eig*. Por ejemplo la matriz diagonal y matriz de paso de la siguiente aplicación lineal:

$$F = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 6 \end{pmatrix}$$

Son las siguientes:

$$P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

Y pueden obtenerse en Matlab con el comando siguiente:

```
>> [P,D] = eig(F)
```

En este caso diagonalizaría F y devolvería la matriz de paso P con los vectores propios y la matriz diagonal D , caso de existir.

0.4. Otros comandos de interés:

Si los vectores $\vec{x} = (x_1, x_2, x_3, \dots, x_n)$ e $\vec{y} = (y_1, y_2, y_3, \dots, y_n)$ contienen un conjunto de datos, su media, varianza se definen como:

$$\mu_x = \frac{1}{n} \sum x_i \qquad \sigma_x^2 = \frac{1}{n} \sum_i (x_i - \mu_x)^2$$

Y su covarianza,

$$\sigma_{xy} = \frac{1}{n} \sum_i (x_i - \mu_x) \cdot (y_i - \mu_y)$$

En Matlab se calculan con los comandos,

$$>> \text{mux} = \text{mean}(x) \quad >> \text{varx} = \text{var}(x) \quad \text{covarxy} = \text{cov}(x,y)$$

Donde guardaría los resultados en las variables *mux*, *varx* y *covarxy* respectivamente. No obstante en los ejercicios que siguen calculamos la varianza y covarianza de un modo diferente.

0.5. Ejercicios.

- Introduce en Matlab los vectores $\vec{x} = (1, 0, 1, 0)$ e $\vec{y} = (-1, 3)$
- Calcula $x' \cdot y$, $x \cdot y$ y $x \times y$
- Introduce en Matlab las matrices:

$$A = \begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix} \quad >> B = \begin{pmatrix} 1 & 0 & 1 \\ 3 & 2 & 0 \\ 0 & 1 & 3 \end{pmatrix}$$

- Calcula: $\vec{x} + \vec{y}$, $5\vec{x} - 3\vec{y}$
- Calcula: $B \cdot B^{-1}$, $B^{-1} \cdot A^{-1}$
- Calcula: $2 \cdot A + B^3 - A^3$, $B^{-1} \cdot |A|$
- Calcula los vectores y valores propios de F .
- Dados los vectores propios $\vec{v}_1 = (1, 0, 1)$, $\vec{v}_2 = (0, 1, 1)$, $\vec{v}_3 = (0, -1, 1)$. Asociados a los valores propios $\lambda_1 = 2$, $\lambda_2 = 2$, $\lambda_3 = -1$. Obtén la matriz asociada a la aplicación lineal en base canónica.

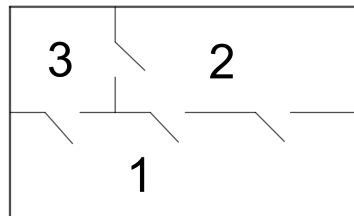
Parte II

Problemas resueltos

Problema 1

Los visitantes aleatorios:

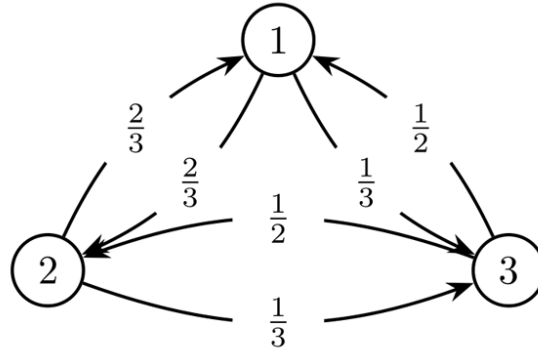
Supongamos que visitan un museo un grupo de visitantes aleatorios. El museo tiene tres salas comunicadas y están comunicadas tal como se indica en la Figura 1. Las salas permanecen cerradas y se abren cierto intervalo de tiempo, digamos 5 minutos. Sabiendo que cada intervalo de tiempo todos los visitantes cambian de ubicación y que los visitantes cuando salen eligen una nueva sala al azar, veamos cuál será su distribución de visitantes al cabo de un intervalo largo de tiempo.



Este tipo de problemas se puede resolver usando una matriz de transición, como los visitantes no tienen ninguna preferencia por las diferentes puertas, en cada paso de tiempo, la probabilidad de transición de una habitación a otra vendrá dada por el número de puertas disponibles, por ejemplo, un visitante en la sala 1 tendrá probabilidad $2/3$ de cambiar a la sala 2 e $1/3$ de hacerlo a la 3. Y así sucesivamente con cada sala. A este tipo de problemas se los denominan cadenas de Markov, conocida la distribución inicial de visitantes en las habitaciones cada paso de tiempo se aplica la matriz de transición para calcular la distribución en el paso siguiente. El diagrama de estados que representa lo descrito es el siguiente:

A partir del diagrama se puede obtener la siguiente matriz de transición:

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



Se observa que los elementos de la matriz corresponden con las probabilidades de transición del diagrama. Dada una población inicial de N visitantes, digamos por ejemplo $N = 10$, inicialmente se distribuyen por las habitaciones, digamos de la forma siguiente

$$\vec{r} = (3, 3, 4)$$

Que indica que hay 3 visitantes en la sala 1, otros 3 en la 2 y 4 visitantes en la 3. La base canónica para este problema sería la siguiente:

$$\begin{cases} \vec{e}_1 = (1, 0, 0) \\ \vec{e}_2 = (0, 1, 0) \\ \vec{e}_3 = (0, 0, 1) \end{cases}$$

Y representa las habitaciones 1, 2 y 3 respectivamente. Los vectores de este espacio representan el número de visitantes en cada sala, es decir:

$$\vec{v}(t) = (X, Y, Z) \in \mathbb{R}$$

Donde t representa la interacción donde estamos. Usando esta notación y la matriz de transición tendríamos el siguiente sistema para la transición del estado 0 al estado 1.

$$\begin{pmatrix} X(1) \\ Y(1) \\ Z(1) \end{pmatrix} = \begin{pmatrix} 0 & 2/3 & 1/2 \\ 2/3 & 0 & 1/2 \\ 1/3 & 1/3 & 0 \end{pmatrix} \begin{pmatrix} X(0) \\ Y(0) \\ Z(0) \end{pmatrix}$$

En notación más condensada el estado 1 se puede escribir en función del 0 usando la matriz de transición:

$$\vec{r}(1) = A\vec{r}(0)$$

Y así sucesivamente con el estado 2, el 3 hasta llegar al estado k

$$\vec{r}(2) = A\vec{r}(1) = A^2\vec{r}(0) \quad \rightarrow \quad \vec{r}(k) = A^k\vec{r}(0)$$

Por lo tanto calculando las potencias de la matriz de transición A^k puedo obtener el número de visitantes en cada sala pasados k intervalos de tiempo. Esto lo podemos lograr de dos maneras:

1.1. Resolución mediante Diagonalización

Calculamos el polinomio característico:

$$p = |A - xI| = \begin{vmatrix} -x & 2/3 & 1/2 \\ 2/3 & -x & 1/2 \\ 1/3 & 1/3 & -x \end{vmatrix} = -x^3 + \frac{14}{18}x + \frac{4}{18} = -18x^3 + 14x + 4 = 0$$

Cuya factorización es la siguiente:

$$(x - 1)(x - 2/3)(x - 1/3) = 0$$

Por lo tanto los valores propios son

$$x_1 = 1 \quad x_2 = -2/3 \quad x_3 = -1/3$$

Los vectores propios pueden obtenerse sustituyendo los valores propios en el polinomio característico y resolviendo el sistema homogéneo para llegar a:

$$\vec{v}_1 = (3, 3, 2) \quad \vec{v}_2 = (-1, 1, 0) \quad \vec{v}_3 = (-1, -1, 2)$$

Por lo tanto la matriz de paso y su inversa son:

$$P = \begin{pmatrix} 3 & -1 & -1 \\ 3 & 1 & -1 \\ 2 & 0 & 2 \end{pmatrix}; P^{-1} = \begin{pmatrix} 2 & 2 & 2 \\ -8 & 8 & 0 \\ -2 & -2 & 6 \end{pmatrix}$$

Siguiendo esta notación, el estado tras k intervalos de tiempo se escribiría:

$$\begin{pmatrix} X(k) \\ Y(k) \\ Z(k) \end{pmatrix} = \begin{pmatrix} 3 & -1 & -1 \\ 3 & 1 & -1 \\ 2 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{-2}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix}^k \cdot \begin{pmatrix} 2 & 2 & 2 \\ -8 & 8 & 0 \\ -2 & -2 & 6 \end{pmatrix} \cdot \begin{pmatrix} X(0) \\ Y(0) \\ Z(0) \end{pmatrix}$$

Si hacemos las operaciones y finalmente tomamos el límite $k \rightarrow \infty$, llegamos a:

$$\begin{pmatrix} X(\infty) \\ Y(\infty) \\ Z(\infty) \end{pmatrix} = \begin{pmatrix} 3/8 & 3/8 & 3/8 \\ 3/8 & 3/8 & 3/8 \\ 1/4 & 1/4 & 1/4 \end{pmatrix} \cdot \begin{pmatrix} X(0) \\ Y(0) \\ Z(0) \end{pmatrix}$$

Escrito por componentes se obtiene:

- $X(\infty) = \frac{3}{8}X(0) + \frac{3}{8}Y(0) + \frac{3}{8}Z(0) \rightarrow \frac{3}{8} \cdot Total$
- $Y(\infty) = \frac{3}{8}X(0) + \frac{3}{8}Y(0) + \frac{3}{8}Z(0) \rightarrow \frac{3}{8} \cdot Total$
- $Z(\infty) = \frac{1}{4}X(0) + \frac{1}{4}Y(0) + \frac{1}{4}Z(0) \rightarrow \frac{1}{4} \cdot Total$

Es decir a largo plazo en la primera sala habrá $3/8$ de los visitantes totales, en la segunda $3/8$ de los visitantes totales y en la tercera $1/4$ de los totales. Independientemente de los visitantes que haya inicialmente. Este es un resultado muy interesante porque nos indica que la distribución de visitantes tiende a un valor estable a largo plazo.

1.2. Resolución numérica.

Procedemos a resolver el problema usando MATLAB y mostraremos que la solución numérica converge hacia la obtenida mediante cálculo. Vamos a suponer tres distribuciones iniciales de visitantes, las siguientes:

$$\begin{cases} \vec{v}_0 = (5, 5, 5) \\ \vec{w}_0 = (0, 8, 7) \\ \vec{r}_0 = (0, 0, 15) \end{cases}$$

En los tres casos tenemos inicialmente 15 visitantes pero distribuidos de diferentes formas.

- En primer lugar introducimos la distribución de visitantes

$$>> v0 = [5, 5, 5]$$

- A continuación la matriz de transición.

$$>> A = [0 \quad 2/3 \quad 1/2; 2/3 \quad 0 \quad 1/2; 1/3 \quad 1/3 \quad 0]$$

- La evolución al paso 2 se obtiene multiplicando ambas, es decir:

$$>> v1 = A * v0'$$

- Se obtiene $v1 = (5,8333, 5,8333, 3,3333)$;
- Donde el apóstrofe indica que el vector $v(1)$ va traspuesto para poder hacer el producto.
- Podríamos repetir sucesivamente este proceso e ir calculando el paso 3, el 4 . . .
- Sin embargo obteniendo las potencias podemos obtener directamente el paso 5 y el 10

- Paso 5

$$>> v5 = A^5 * v0'$$

- Obtenido $v5 = (5,6276, 5,6276, 3,7449)$;
- Y finalmente el paso 10.

$$>> v10 = A^{10} * v0'$$

- Obtenido $v10 = (5,6250, 5,6250, 3,7500)$;

El proceso es exactamente el mismo para los vectores iniciales \vec{w} y \vec{r} . La siguiente tabla muestra los resultados.

-	Paso0	Paso 1	Paso 5	Paso 10
\vec{v}	(5, 5, 5)	(5,8333, 5,8333, 3,3333)	(5,6276, 5,6276, 3,7449)	(5,6250, 5,6250, 3,7500)
\vec{w}	(0, 8, 7)	(8,8333, 3,5000, 2,6667)	(6,1584, 5,1049, 3,7366)	(5,5556, 5,6943, 3,7501)
\vec{r}	(0, 0, 15)	(5, 5, 5)	(5,6481, 5,6481, 3,7037)	(5,6249, 5,6249, 3,7502)

Se observa como el vector \vec{r} reproduce el vector \vec{v} en el paso 2 y a partir de ese momento seguirá si misma evolución pero con cierto desfase. La resolución por diagonalización del problema nos dice que la distribución a largo plazo de visitantes sólo depende del número total de éstos no de como estén distribuidos inicialmente. En concreto para $N = 15$ visitantes tendremos la siguiente.

$$v_{\infty} = \left(\frac{3}{8}N, \frac{3}{8}N, \frac{1}{4}N\right) = (5,6250, 5,6250, 3,7500)$$

Y vemos que efectivamente los tres casos se aproximan sucesivamente al resultado esperado.

1.3. Resolución numérica usando diagonalización.

También se puede resolver rápidamente usando Matlab y diagonalización. Obtenemos la matriz diagonal D y la matriz de paso P usando el comando:

$$>> [P,D] = eig(A)$$

Obteniendo

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{2}{3} & 0 \\ 0 & 0 & -\frac{1}{3} \end{pmatrix} ; \quad P = \begin{pmatrix} -0,6396 & -0,7071 & 0,8165 \\ -0,6396 & -0,7071 & -0,4082 \\ -0,4264 & 0,0000 & 0,8165 \end{pmatrix}$$

Donde se observa que las columnas de la matriz de paso son proporcionales a las obtenidas mediante cálculo. Para obtener la distribución de rosas en cualquier instante es suficiente con calcular la potencia como nos indica la teoría de diagonalización, por ejemplo la décima generación:

$$\gg v(10) = P * D^{10} * inv(P) * v'$$

Para obtener:

$$v(10) = (5,6250, 5,6250, 3,7500)$$

Resultado equivalente al anterior.

Problema 2

Herencia Mendeliana.

Un vivero tiene una gran población de rosas de tres colores, rojo, rosa y blanco. El color de una rosa es rojo si tiene genotipo AA, rosa si es Aa y blanco si es aa. Dicho vivero inicia un programa de cultivo en el que, todas las plantas son fecundadas con ejemplares tipo AA, es decir rojos. Se van obteniendo sucesivas generaciones. Dado que conocemos la distribución inicial de colores, queremos obtener como evoluciona dicha distribución y eventualmente, obtener la distribución a largo plazo del color de nuestras rosas.



Estamos ante un problema similar al anterior, entre una generación y la siguiente debemos utilizar una matriz de transición, en este caso el genotipo de la generación siguiente en términos de la anterior vendrá dado por la tabla siguiente.

		PROGENITORES					
		AAxAA	AAxAa	AAxaa	AaxAa	Aaxaa	aaxaa
PROGENIE	AA	1	0.5	0	0.25	0	0
	Aa	0	0.5	1	0.5	0.5	0
	aa	0	0	0	0.25	0.5	1

El enunciado del problema indica que todas las rosas del vivero se fecundan con rosas rojas, por tanto con genotipo (AA). Este dato nos permite simplificar la tabla de herencia autosómica y, dado que una de las rosas de cada cruce es AA, quedarnos únicamente con las tres primeras columnas. Dichas columnas determinan la matriz de transición, la siguiente:

$$A = \begin{pmatrix} 1 & 0,5 & 0 \\ 0 & 0,5 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Si consideramos una población inicial de rosas, digamos $N = 100$, un vector de este espacio nos indicará cuantas de cada color tenemos, por ejemplo

$$\vec{v}_0 = (35, 35, 30)$$

Me indicaría que hay 35 rosas rojas, otras 35 rosas rosas (valga la redundancia) y 30 blancas. Estamos por tanto ante un espacio vectorial de dimensión 3, una de cuyas posibles bases es la canónica, es decir.

$$\begin{cases} \vec{e}_1 = (1, 0, 0) \\ \vec{e}_2 = (0, 1, 0) \\ \vec{e}_3 = (0, 0, 1) \end{cases}$$

Cuyos vectores representarían los colores rojo, rosa y blanco para nuestro problema. Un vector general que represente cualquier distribución inicial de colores se escribiría:

$$\vec{v}(t) = (X, Y, Z) \in \mathbb{R}$$

Donde t representa la interacción donde estamos. Usando esta notación y la matriz de transición tendríamos el siguiente sistema para la transición del estado 0 al estado 1.

$$\begin{pmatrix} X(1) \\ Y(1) \\ Z(1) \end{pmatrix} = \begin{pmatrix} 1 & 0,5 & 0 \\ 0 & 0,5 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} X(0) \\ Y(0) \\ Z(0) \end{pmatrix}$$

Es notación más condensada el estado 1 se puede escribir en función del 0 usando la matriz de transición:

$$\vec{v}(1) = A\vec{v}(0)$$

Y así sucesivamente con el estado 2, el 3 hasta llegar al estado k

$$\vec{v}(2) = A\vec{v}(1) = A^2\vec{v}(0) \quad \rightarrow \quad \vec{v}(k) = A^k\vec{v}(0)$$

Por lo tanto calculando las potencias de la matriz de transición A^k puedo obtener el número de visitantes en cada sala pasados k intervalos de tiempo. En primer lugar vamos a resolver el problema mediante diagonalización, a continuación numéricamente usando MATLAB, acabaremos comparando ambos resultados.

2.1. Resolución mediante Diagonalización

Calculamos el polinomio característico:

$$p = |A - xI| = \begin{vmatrix} 1-x & 0,5 & 0 \\ 0 & 0,5-x & 0 \\ 0 & 0 & -x \end{vmatrix} = -x \cdot (1-x)(0,5-x) = 0$$

Por lo tanto los valores propios de esta aplicación son 0, 1 y 1/2.

$$(x-1)(x-2/3)(x-1/3) = 0$$

Por lo tanto los valores propios son

$$x_1 = 1 \quad x_2 = -2/3 \quad x_3 = -1/3$$

Los vectores propios pueden obtenerse sustituyendo los valores propios en el polinomio característico y resolviendo el sistema homogéneo para llegar a:

$$\vec{v}_1 = (1, 0, 0) \quad \vec{v}_2 = (1, 1, 0) \quad \vec{v}_3 = (1, -2, 1)$$

Por lo tanto la matriz de paso y su inversa son:

$$P = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{pmatrix}; P^{-1} = \begin{pmatrix} 1 & -1 & -3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$$

Siguiendo esta notación, el estado tras k intervalos de tiempo se escribiría:

$$\begin{pmatrix} X(k) \\ Y(k) \\ Z(k) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix}^k \cdot \begin{pmatrix} 1 & -1 & -3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} X(0) \\ Y(0) \\ Z(0) \end{pmatrix}$$

Haciendo las operaciones se obtiene:

$$\begin{pmatrix} X(k) \\ Y(k) \\ Z(k) \end{pmatrix} = \begin{pmatrix} 1 & 1 - (1/2)^k & 1 - (1/2)^{k-1} \\ 0 & (1/2)^k & (1/2)^{k-1} \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} X(0) \\ Y(0) \\ Z(0) \end{pmatrix}$$

Y finalmente tomamos el límite $k \rightarrow \infty$, para llegar a:

$$\begin{pmatrix} X(\infty) \\ Y(\infty) \\ Z(\infty) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} X(0) \\ Y(0) \\ Z(0) \end{pmatrix}$$

Por lo tanto cuando el número de generaciones crece el color de las rosas tiende a ser únicamente rojo.

$$\vec{v}_\infty = (X(0) + Y(0) + Z(0), 0, 0) = (N, 0, 0)$$

2.2. Resolución numérica.

Procedemos a resolver el problema usando MATLAB y mostraremos que la solución numérica converge hacia la obtenida mediante cálculo. Vamos a suponer tres distribuciones iniciales de visitantes, las siguientes:

$$\begin{cases} \vec{v}_0 = (35, 35, 30) \\ \vec{w}_0 = (5, 50, 45) \\ \vec{r}_0 = (0, 1, 99) \end{cases}$$

En los tres casos tenemos inicialmente 15 visitantes pero distribuidos de diferentes formas.

- En primer lugar introducimos la distribución de rosas

$$\gg v0 = [353530]$$

- A continuación la matriz de transición.

$$\gg A = [10,50;00,51;000]$$

- La evolución al paso 2 se obtiene multiplicando ambas, es decir:

$$\gg v1 = A * v0'$$

- Se obtiene $v1 = (52,5,47,5,0)$;
- Donde el apóstrofe indica que el vector $v(1)$ va traspuesto para poder hacer el producto.
- Podríamos repetir sucesivamente este proceso e ir calculando el paso 2, el 4 . . .
- Sin embargo obteniendo las potencias podemos obtener directamente el paso 5 y el 10
- Paso 2

$$\gg v2 = A^2 * v0'$$

- Obteniedo $v2 = (70,25,23,75,0)$;
- Y finalmente el paso 10.

$$\gg v10 = A^{10} * v0'$$

- Obtenido $v_{10} = (99,9,0,1,0)$;

El proceso es exactamente el mismo para los vectores iniciales \vec{w} y \vec{r} . La siguiente tabla muestra los resultados.

-	Paso1	Paso 2	Paso 5	Paso 10
\vec{v}	(35, 35, 30)	(52,5, 47,5, 0)	(70,25, 23,75, 0)	(99,9, 0,1, 0)
\vec{w}	(5, 50, 45)	(30, 70, 0)	(65, 35, 0)	(99,86, 0,14, 0)
\vec{r}	(0, 1, 99)	(0,5, 99,5, 0)	(50,25, 49,75, 0)	(99,81, 0,19, 0)

Vemos como ya en la primera generación desaparecen las rosas blancas (se puede deducir observando la matriz de transición), y transcurridas 10 generaciones más de un 99 % son rojas, independientemente de la distribución de colores original. Este es el resultado predicho por la diagonalización.

2.2.1. Resolución numérica usando diagonalización.

También se puede resolver rápidamente usando Matlab y diagonalización. Obtenemos la matriz diagonal D y la matriz de paso P usando el comando:

$$\gg [P,D] = eig(A)$$

Obteniendo

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} ; \quad P = \begin{pmatrix} 1 & -0,7071 & ,4082 \\ 0 & 0,7071 & 0,8165 \\ 0 & 0 & 0,4082 \end{pmatrix}$$

Donde se observa que las columnas de la matriz de paso son proporcionales a las obtenidas mediante cálculo. Para obtener la distribución de rosas en cualquier instante es suficiente con calcular la potencia como nos indica la teoría de diagonalización, por ejemplo la décima generación:

$$\gg v_{10} = P * D^{10} * inv(P) * v'$$

Para obtener:

$$v_{10} = (99,8057, 0,1943, 0)$$

Resultado equivalente al anterior.

Problema 3

Muestreo de aves cantoras.

Hemos completado un muestreo de 25 especies de aves cantoras tomando datos de su longitud, envergadura y peso. Nos preguntamos sobre si existe alguna relación entre ellas. Para lograrlo vamos a recurrir al cálculo matricial y a la diagonalización completando lo que se conoce como un Análisis de Componentes Principales (PCA). Dicho análisis nos va a permitir averiguar cuál o cuáles son las relaciones ocultas entre estas variables y cuál es el peso relativo de cada una. Para ello necesitaremos también ciertas nociones de estadística.



3.1. Conceptos de estadística:

Supongamos un conjunto de datos de tres variables, por ejemplo longitud (x), envergadura (y) y peso (z) de un conjunto de aves, en este caso 25 especies de aves cantoras. Para cada una de las variables podemos calcular su media y su varianza,

$$\mu_x = \frac{1}{n} \sum x_i \qquad \sigma_x^2 = \frac{1}{n} \sum (x_i - \mu_x)^2$$

Para cada una de las variables, es decir, μ_x , μ_y , μ_z , σ_x^2 , σ_y^2 , σ_z^2 . Y definimos una variable adicional, **la covarianza** o varianza conjunta, definida como,

$$\sigma_{xy} = \frac{1}{n} \sum_i (x_i - \mu_x) \cdot (y_i - \mu_y)$$

Nos permite averiguar si hay alguna relación estadística entre dos variables. Como tenemos tres variables las covarianzas serán,

$$\sigma_{xy}, \sigma_{xz}, \sigma_{yz}$$

ya que son magnitudes simétricas. Usando varianzas y covarianzas construimos **la matriz de covarianza**, viene dada por:

$$M = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_z^2 \end{pmatrix}$$

Y se puede obtener rápidamente a partir de los datos, como veremos.

3.2. Análisis de componentes principales:

Una manera rápida y eficiente de obtener la matriz de covarianza es organizar los datos de nuestras variables en forma de matriz y multiplicarla por su traspuesta. Si nuestra matriz de datos es la siguiente:

$$A = \begin{pmatrix} (x_1 - \mu_x) & (x_2 - \mu_x) & \cdots & (x_n - \mu_x) \\ (y_1 - \mu_y) & (y_2 - \mu_y) & \cdots & (y_n - \mu_y) \\ (z_1 - \mu_z) & (z_2 - \mu_z) & \cdots & (z_n - \mu_z) \end{pmatrix}$$

Donde a cada dato le restamos la media (para que cada variable parta de media cero). La matriz de covarianza se puede obtener operando como sigue:

$$Cov = \frac{1}{N} (A \cdot A^t)$$

Pues:

$$\begin{aligned} Cov = \frac{1}{N} (A \cdot A^t) &= \frac{1}{N} \begin{pmatrix} (x_1 - \mu_x) & (x_2 - \mu_x) & \cdots & (x_n - \mu_x) \\ (y_1 - \mu_y) & (y_2 - \mu_y) & \cdots & (y_n - \mu_y) \\ (z_1 - \mu_z) & (z_2 - \mu_z) & \cdots & (z_n - \mu_z) \end{pmatrix} \cdot \begin{pmatrix} (x_1 - \mu_x) & (y_1 - \mu_y) & (z_1 - \mu_z) \\ (x_2 - \mu_x) & (y_2 - \mu_y) & (z_2 - \mu_z) \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ (x_n - \mu_x) & (y_n - \mu_y) & (z_n - \mu_z) \end{pmatrix} = \cdots \\ \cdots &= \frac{1}{N} \begin{pmatrix} \sum (x_i - \mu_x)^2 & \sum (x_i - \mu_x) \cdot (y_i - \mu_y) & \sum (x_i - \mu_x) \cdot (z_i - \mu_z) \\ \sum_i (x_i - \mu_x) \cdot (y_i - \mu_y) & \sum (y_i - \mu_y)^2 & \sum_i (y_i - \mu_y) \cdot (z_i - \mu_z) \\ \sum_i (x_i - \mu_x) \cdot (z_i - \mu_z) & \sum_i (y_i - \mu_y) \cdot (z_i - \mu_z) & \sum (z_i - \mu_z)^2 \end{pmatrix} \end{aligned}$$

Que son las definiciones de varianza y covarianza. Usando una notación más condensada la matriz de covarianza será:

$$Cov = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_z^2 \end{pmatrix}$$

Donde las σ_i^2 representan las varianzas y las σ_{ij} las covarianzas. Esta matriz será siempre simétrica porque se cumple que $\sigma_{ij} = \sigma_{ji}$. Si diagonalizamos obtendremos un cambio de base similar al siguiente.

$$Cov = P \cdot D \cdot P^{-1} = \begin{pmatrix} v1_x & v2_x & v3_x \\ v1_y & v2_y & v3_y \\ v1_z & v2_z & v3_z \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \cdot \begin{pmatrix} v1_x & v2_x & v3_x \\ v1_y & v2_y & v3_y \\ v1_z & v2_z & v3_z \end{pmatrix}^{-1}$$

Si hemos ordenado los valores propios tal que $\lambda_1 > \lambda_2 > \lambda_3$, y teniendo en cuenta que la matriz de partida es una matriz de covarianza, se puede demostrar que cada uno de los **vectores propios** permite interpretar la variación de los datos de entrada a razón del *peso* de su **valor propio**, es decir, \vec{v}_1 permite interpretar el $\frac{\lambda_1}{\sum \lambda_i}$ (en tanto por uno) de la variación de los datos, \vec{v}_2 el $\frac{\lambda_2}{\sum \lambda_i}$ de los datos y así sucesivamente.

O dicho de otro modo, en la base dada por P , las covarianzas son cero y los valores propios corresponden a las varianzas y están ordenados, el primer vector propio \vec{v}_1 corresponde a la mayor varianza λ_1 y así sucesivamente, permitiendo por tanto interpretar la variación de los datos.

3.3. Caso práctico:

Procedemos ahora a analizar los datos de nuestro muestreo. En primer lugar debemos introducir los datos en Matlab, suponiendo que disponemos de ellos en formato Excel, el proceso de introducción en forma matricial se puede hacer mediante un copiado y pegado. Los datos son los siguientes:

ESPECIES	LONGITUD (cm)	ENVERGADURA(cm)	PESO (g)
Ruiseñor común	17.75	25	25
Jilguero europeo	12.5	21.5	19
Pardillo común	12.75	24	19
Verdecillo	11.75	22	14.5
Petirrojo europeo	13.75	21	19
Carbonero común	11.25	18.5	9.5
Herrerillo común	11.5	17.5	9.5
Pinzón vulgar	15	25.5	30
Curruca cabecinegra	11.75	21	12
Mosquitero común	9.75	15	6
Bisbita común	14.5	23	21.5
Lavandera blanca	17.25	28	24.5
Gorrión común	15	23.5	31.5
Canario	12	22	17
Verderón común	12.25	21.5	17
Alondra común	17.25	34.5	65
Currucas capirotadas	13.5	22.5	14
Mosquitero ibérico	10.25	15	6.5
Ruiseñor bastardo	17.75	26.5	27.5
Tarabilla común	12.75	19	15
Agateador europeo	12.75	23	8.5
Carricero común	12.5	20	12
Carricerín común	10.25	18	7.5
Collalba gris	14	22	16.5
Currucas mosquiteras	13.25	21.5	14.5

Los introducimos en Matlab mediante la matriz siguiente:

$$Data = \begin{pmatrix} 17,75 & 12,5 & \dots & \dots & 13,25 \\ 25 & 21,5 & \dots & \dots & 21,5 \\ 25 & 19 & \dots & \dots & 14,5 \end{pmatrix}$$

Donde hemos omitido los datos centrales por simplicidad. En primer lugar sacamos las medias y las varianzas.

$$\begin{cases} \mu_x = 13,32 \text{ cm} & \dots & \sigma_x^2 = 5,2996 \text{ cm}^2 \\ \mu_y = 22,04 \text{ cm} & \dots & \sigma_y^2 = 16,852 \text{ cm}^2 \\ \mu_z = 18,48 \text{ gr} & \dots & \sigma_z^2 = 144,72 \text{ gr}^2 \end{cases}$$

Restamos la media de cada conjunto de datos de la variable *Data* para sacar la matriz *A*

$$A = \begin{pmatrix} 4,43 & -0,82 & \dots & \dots & -0,07 \\ 2,96 & -0,54 & \dots & \dots & -0,54 \\ 6,52 & 0,52 & \dots & \dots & -3,98 \end{pmatrix}$$

A partir de la cual podemos sacar la matriz de covarianza,

$$Cov = \frac{1}{n-1}(A * A')$$

Para obtener:

$$Cov = \begin{pmatrix} 5,2996 & 18,0648 & 20,861 \\ 8,0648 & 16,852 & 43,678 \\ 20,861 & 43,678 & 144,72 \end{pmatrix}$$

Diagonalizando la matriz de covarianza obtenemos los vectores y valores propios.

$$P = \begin{pmatrix} 0,82147 & 0,55242 & 0,14151 \\ -0,56776 & 0,76907 & 0,29356 \\ 0,053342 & -0,32149 & 0,94541 \end{pmatrix} \quad D = \begin{pmatrix} 1,0802 & 0 & 0 \\ 0 & 4,3869 & 0 \\ 0 & 0 & 161,4 \end{pmatrix}$$

Observamos que $\lambda_3 \sim 161,4$ corresponde con $\sim 96\%$ de la varianza total y su vector propio es $\vec{v}_3 = (0,14,0,29,0,94)$, el segundo valor propio $\lambda_2 \sim 4,4$ corresponde a $\sim 2,5\%$ de la varianza total y finalmente $\lambda_1 \sim 1,08$ únicamente corresponde a un $\sim 0,6\%$ de la varianza. El primer vector propio indica la dirección donde apunta la gran mayoría de la varianza de los datos, con clara predominancia para el peso, aunque se observa cierta correlación con longitud y envergadura.

Problema 4

Estadística de hojas de árbol

La gran diversidad de formas y tamaños de las hojas de los árboles supone un reto. Los parámetros que se utilizan para caracterizar una hoja de una planta pueden variar según la especie de planta y el propósito de la investigación, pero aquí hay algunos parámetros comunes que se utilizan para describir la estructura y las propiedades de las hojas, serían los siguientes:

- Área foliar
- Longitud y ancho de la hoja
- Espesor de la hoja
- Peso de la hoja
- Concentración de clorofila.
- Área específica de la hoja (área foliar por unidad de peso seco)
- Densidad de la hoja
- Índice de área foliar (IAF): es una medida de la cantidad total de hojas en una planta y se calcula dividiendo el área foliar total por el área del suelo ocupada por la planta.



En este problema vamos a trabajar con tres de estas variables, en concreto la longitud, la anchura y la concentración de clorofila de un conjunto de especies.

4.1. Conceptos de estadística:

Supongamos un conjunto de datos de tres variables, por ejemplo longitud (x), envergadura (y) y peso (z) de un conjunto de aves, en este caso 25 especies de aves cantoras. Para cada una de las variables podemos calcular su media y su varianza,

$$\mu_x = \frac{1}{n} \sum x_i \qquad \sigma_x^2 = \frac{1}{n} \sum_i (x_i - \mu_x)^2$$

Para cada una de las variables, es decir, μ_x , μ_y , μ_z , σ_x^2 , σ_y^2 , σ_z^2 . Y definimos una variable adicional, **la covarianza** o varianza conjunta, definida como,

$$\sigma_{xy} = \frac{1}{n} \sum_i (x_i - \mu_x) \cdot (y_i - \mu_y)$$

Nos permite averiguar si hay alguna relación estadística entre dos variables. Como tenemos tres variables las covarianzas serán,

$$\sigma_{xy}, \sigma_{xz}, \sigma_{yz}$$

ya que son magnitudes simétricas. Usando varianzas y covarianzas construimos **la matriz de covarianza**, viene dada por:

$$Cov = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_z^2 \end{pmatrix}$$

Matriz que puede calcularse fácil y rápidamente a partir de los datos, como veremos.

4.2. Análisis de componentes principales:

Una manera rápida y eficiente de obtener la matriz de covarianza es organizar los datos de nuestras variables en forma de matriz y multiplicarla por su traspuesta. Si nuestra matriz de datos es la siguiente:

$$A = \begin{pmatrix} (x_1 - \mu_x) & (x_2 - \mu_x) & \cdots & (x_n - \mu_x) \\ (y_1 - \mu_y) & (y_2 - \mu_y) & \cdots & (y_n - \mu_y) \\ (z_1 - \mu_z) & (z_2 - \mu_z) & \cdots & (z_n - \mu_z) \end{pmatrix}$$

Donde a cada dato le restamos la media (para que cada variable parta de media cero). La matriz de covarianza se puede obtener operando como sigue:

$$Cov = \frac{1}{N}(A \cdot A^t)$$

Pues:

$$\begin{aligned} Cov &= \frac{1}{N}(A \cdot A^t) = \frac{1}{N} \begin{pmatrix} (x_1 - \mu_x) & (x_2 - \mu_x) & \cdots & (x_n - \mu_x) \\ (y_1 - \mu_y) & (y_2 - \mu_y) & \cdots & (y_n - \mu_y) \\ (z_1 - \mu_z) & (z_2 - \mu_z) & \cdots & (z_n - \mu_z) \end{pmatrix} \cdot \begin{pmatrix} (x_1 - \mu_x) & (y_1 - \mu_y) & (z_1 - \mu_z) \\ (x_2 - \mu_x) & (y_2 - \mu_y) & (z_2 - \mu_z) \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ (x_n - \mu_x) & (y_n - \mu_y) & (z_n - \mu_z) \end{pmatrix} = \cdots \\ &\cdots = \frac{1}{N} \begin{pmatrix} \sum (x_i - \mu_x)^2 & \sum_i (x_i - \mu_x) \cdot (y_i - \mu_y) & \sum_i (x_i - \mu_x) \cdot (z_i - \mu_z) \\ \sum_i (x_i - \mu_x) \cdot (y_i - \mu_y) & \sum (y_i - \mu_y)^2 & \sum_i (y_i - \mu_y) \cdot (z_i - \mu_z) \\ \sum_i (x_i - \mu_x) \cdot (z_i - \mu_z) & \sum_i (y_i - \mu_y) \cdot (z_i - \mu_z) & \sum (z_i - \mu_z)^2 \end{pmatrix} \end{aligned}$$

Que son las definiciones de varianza y covarianza. Usando una notación más condensada la matriz de covarianza será:

$$Cov = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_z^2 \end{pmatrix}$$

Donde las σ_i^2 representan las varianzas y las σ_{ij} las covarianzas. Esta matriz será siempre simétrica porque se cumple que $\sigma_{ij} = \sigma_{ji}$. Si diagonalizamos obtendremos un cambio de base similar al siguiente.

$$Cov = P \cdot D \cdot P^{-1} = \begin{pmatrix} v1_x & v2_x & v3_x \\ v1_y & v2_y & v3_y \\ v1_z & v2_z & v3_z \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \begin{pmatrix} v1_x & v2_x & v3_x \\ v1_y & v2_y & v3_y \\ v1_z & v2_z & v3_z \end{pmatrix}^{-1}$$

Si hemos ordenado los valores propios tal que $\lambda_1 > \lambda_2 > \lambda_3$, y teniendo en cuenta que la matriz de partida es una matriz de covarianza, se puede demostrar que cada uno de los **vectores propios** permite interpretar la variación de los datos de entrada a razón del *peso* de su **valor propio**, es decir, \vec{v}_1 permite interpretar el $\frac{\lambda_1}{\sum \lambda_i}$ (en tanto por uno) de la variación de los datos, \vec{v}_2 el $\frac{\lambda_2}{\sum \lambda_i}$ de los datos y así sucesivamente.

O dicho de otro modo, en la base dada por P , las covarianzas son cero y los valores propios corresponden a las varianzas y están ordenados, el primer vector propio \vec{v}_1 corresponde a la mayor varianza λ_1 y así sucesivamente, permitiendo por tanto interpretar la variación de los datos.

4.3. Caso práctico.

Los datos obtenidos son los siguientes:

		VARIABLES		
		LONGITUD	ANCHURA	CLOROFILA
ESPECIES	Arce	12.5	9.5	0.8
	Roble	10	12.5	0.5
	Olmo	8	12.5	0.9
	Abedul	6	7.5	0.3
	Sauce	6	10	0.8
	Haya	10	14	0.9
	Chopo	8	17.5	0.3
	Nogal	7.5	17.5	0.9
	Encina	4.5	7.5	0.9
	Olivo	6	9	0.9
	Ciprés	4	10	0.2
	Granado	6	10	0.6
	Manzano	6	10	0.8
	Melojo	10	12.5	0.5
	Rebollo	8	15	0.4
	Castaño	10	15	0.7
Plátano	12.5	17.5	0.5	

Los introducimos en Matlab mediante la matriz siguiente:

$$Data = \begin{pmatrix} 12,5 & 10,0 & \dots & \dots & 12,5 \\ 9,5 & 12,5 & \dots & \dots & 17,5 \\ 0,8 & 0,5 & \dots & \dots & 0,5 \end{pmatrix}$$

Donde hemos omitido los datos centrales por simplicidad. En primer lugar sacamos las medias y las varianzas.

$$\begin{cases} \mu_x = 7,94 & cm & \dots & \sigma_x^2 = 6,59 & cm^2 \\ \mu_y = 12,21 & cm & \dots & \sigma_y^2 = 11,53 & cm^2 \\ \mu_z = 0,64 & -- & \dots & \sigma_z^2 = 0,0601 & -- \end{cases}$$

Restamos la media de cada conjunto de datos de la variable *Data* para sacar la matriz *A*

$$A = \begin{pmatrix} 4,56 & 2,06 & \dots & \dots & 4,56 \\ -2,71 & 0,29 & \dots & \dots & 5,25 \\ 0,16 & -0,14 & \dots & \dots & -0,14 \end{pmatrix}$$

A partir de la cual podemos sacar la matriz de covarianza,

$$Cov = \frac{1}{n-1}(A * A')$$

Para obtener:

$$Cov = \begin{pmatrix} 6,5588 & 4,6066 & 0,0307 \\ 4,6066 & 11,5331 & -0,1090 \\ 0,0307 & -0,1090 & 0,0601 \end{pmatrix}$$

Diagonalizando la matriz de covarianza obtenemos los vectores y valores propios.

$$P = \begin{pmatrix} 0,5123 & 0,8587 & -0,0160 \\ 0,8588 & -0,5120 & 0,0159 \\ -0,0055 & 0,0219 & 0,9997 \end{pmatrix} \quad D = \begin{pmatrix} 14,28 & 0 & 0 \\ 0 & 3,81 & 0 \\ 0 & 0 & 0,06 \end{pmatrix}$$

El primer autovalor $\lambda_1 = 14,28$ explica un $\sim 78,7\%$ de la varianza el vector propio apunta en el plano XY, es decir, sin participación de la concentración de clorofila. El segundo autovalor $\lambda_2 = 3,81$ explica un $\sim 21\%$ de la varianza y el autovector también apunta en dicho plano, siendo además aproximadamente perpendicular al primero. Y el tercer valor propio es residual con un $\sim 0,32\%$ de la varianza con vector propio aproximadamente sobre el eje Z, es decir, en la dirección de la concentración de clorofila, indicando que esta variable es independiente de las dos anteriores.

Parte III

Enunciados para los alumnos

0.4. Instrucciones generales.

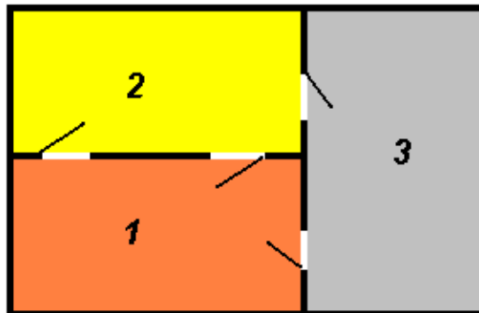
Lee atentamente estas instrucciones para que tu ejercicio pueda ser evaluado.

- Todos los comando necesarios para resolver el problema deben guardarse en un archivo *.m (un script), de tal manera que sea ejecutable y no de errores.
- Pueden añadirse comentarios para explicar los comandos (los comentarios empiezan por el caracter %)
- **No deben añadirse** los resultados que saca el programa por consola. Éstos se pueden reproducir ejecutando el comando.
- Los errores de código y líneas que no puedan ejecutarse en Matlab restarán en la puntuación final del ejercicio.
- Debe nombrarse el ejercicio como *nombre_apellido_tic1.m*, obviamente usando vuestro nombre y apellido.
- Dichos archivos deben subirse al moodle en la entrega preparada al efecto.
- No está permitido compartir código ni datos entre alumnos. Estará penalizado.
- El profesor podrá pedir al alumno que explique las líneas de código que sean necesarias o repita el ejercicio para demostrar el dominio del problema.

Problema 1

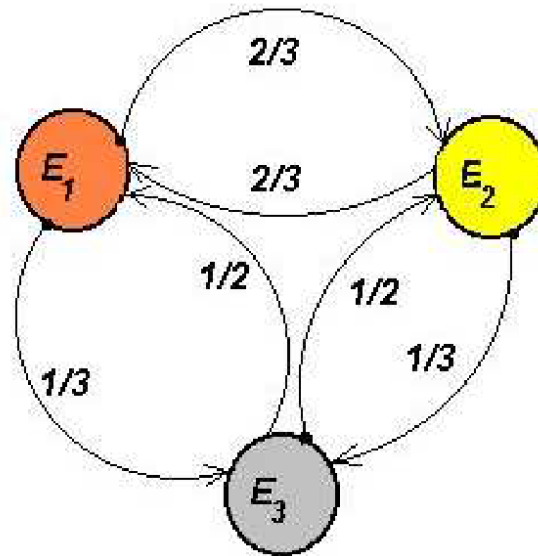
Los visitantes aleatorios:

Supongamos que visitan un museo un grupo de visitantes aleatorios. El museo tiene tres salas comunicadas y están comunicadas tal como se indica en la Figura 1. Las salas permanecen cerradas y se abren cierto intervalo de tiempo, digamos 5 minutos. Sabiendo que cada intervalo de tiempo todos los visitantes cambian de ubicación y que los visitantes cuando salen eligen una nueva sala al azar, veamos cuál será su distribución de visitantes al cabo de un intervalo largo de tiempo.



Este tipo de problemas se puede resolver usando una matriz de transición, como los visitantes no tienen ninguna preferencia por las diferentes puertas, en cada paso de tiempo, la probabilidad de transición de una habitación a otra vendrá dada por el número de puertas disponibles, por ejemplo, un visitante en la sala 1 tendrá probabilidad $2/3$ de cambiar a la sala 2 e $1/3$ de hacerlo a la 3. Y así sucesivamente con cada sala. A este tipo de problemas se los denominan cadenas de Markov, conocida la distribución inicial de visitantes en las habitaciones cada paso de tiempo se aplica la matriz de transición para calcular la distribución en el paso siguiente. El diagrama de estados que representa lo descrito es el siguiente:

A partir del diagrama se puede obtener la siguiente matriz de transición:



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

Y la transición entre estados vendría dada por:

$$\begin{pmatrix} X(n) \\ Y(n) \\ Z(n) \end{pmatrix} = \begin{pmatrix} 0 & 2/3 & 1/2 \\ 2/3 & 0 & 1/2 \\ 1/3 & 1/3 & 0 \end{pmatrix} \begin{pmatrix} X(n-1) \\ Y(n-1) \\ Z(n-1) \end{pmatrix}$$

1.1. Preguntas.

Se pide lo siguiente

- Obtener, usando Matlab, la distribución de visitantes por habitación tras la primera interacción, la quinta y la décima interacción para para las siguientes distribuciones iniciales.

$$\begin{cases} \vec{r}_0 = (5, 5, 5) \\ \vec{r}_1 = (0, 8, 7) \\ \vec{r}_2 = (0, 0, 15) \end{cases}$$

- Calcula la distribución de visitantes para la octava interacción usando diagonalización y el siguiente vector inicial

$$\vec{r}_0 = (0, 5, 10)$$

- Comparar y comentar los resultados con el resultado teórico. ¿Se cumple lo esperado? ¿Por qué?
- ¿Por qué el resultado final es independiente de la distribución inicial? ¿Cuál sería la distribución inicial de 15 visitantes que más tardaría en alcanzar ese estado?

Problema 2

Herencia Mendeliana.

Un vivero tiene una gran población de rosas de tres colores, rojo, rosa y blanco. El color de una rosa es rojo si tiene genotipo AA, rosa si es Aa y blanco si es aa. Dicho vivero inicia un programa de cultivo en el que, todas las plantas son fecundadas con ejemplares tipo AA, es decir rojos. Se van obteniendo sucesivas generaciones. Dado que conocemos la distribución inicial de colores, queremos obtener como evoluciona dicha distribución y eventualmente, obtener la distribución a largo plazo del color de nuestras rosas.



Estamos ante un problema similar al anterior, entre una generación y la siguiente debemos utilizar una matriz de transición, en este caso el genotipo de la generación siguiente en términos de la anterior vendrá dado por la tabla siguiente.

		PROGENITORES					
		AAxAA	AAxAa	AAxaa	AaxAa	Aaxaa	aaxaa
PROGENIE	AA	1	0.5	0	0.25	0	0
	Aa	0	0.5	1	0.5	0.5	0
	aa	0	0	0	0.25	0.5	1

El enunciado del problema indica que todas las rosas del vivero se fecundan con rosas rojas, por tanto con genotipo (AA). Este dato nos permite simplificar la tabla de herencia autosómica y, dado que una de las rosas de cada cruce es AA, quedarnos únicamente con las tres primeras columnas. Dichas columnas determinan la matriz de transición, la siguiente:

$$A = \begin{pmatrix} 1 & 0,5 & 0 \\ 0 & 0,5 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Si la distribución de colores cuando el número de generaciones tiende a infinito viene dado por el sistema:

$$\begin{pmatrix} X(\infty) \\ Y(\infty) \\ Z(\infty) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} X(0) \\ Y(0) \\ Z(0) \end{pmatrix}$$

2.1. Preguntas.

Se pide lo siguiente:

- Obtener la distribución de colores de las rosas para la segunda, cuarta y octava generación de rosas usando Matlab, para los siguientes vectores iniciales de rosas.

$$\begin{cases} \vec{v}_0 = (35, 35, 30) \\ \vec{w}_0 = (5, 50, 45) \\ \vec{r}_0 = (0, 1, 99) \end{cases}$$

- Calcula la distribución de colores de las rosas para la sexta generación usando diagonalización y el siguiente vector inicial

$$\vec{r}_0 = (0, 30, 70)$$

- Comparar y comentar los resultados con el resultado teórico. ¿Se cumple lo esperado? ¿Por qué?
- ¿Cómo tendría que ser la matriz de transición para que la distribución de rosas $\vec{r}_\infty = (0, 100, 0)$? ¿Y $\vec{r}_\infty = (40, 60, 0)$?

Problema 3

Muestreo de aves cantoras.

Hemos completado un muestreo de 25 especies de aves cantoras tomando datos de su longitud, envergadura y peso. Nos preguntamos sobre si existe alguna relación entre ellas. Para lograrlo vamos a recurrir al cálculo matricial y a la diagonalización completando lo que se conoce como un Análisis de Componentes Principales (PCA). Dicho análisis nos va a permitir averiguar cuál o cuáles son las relaciones ocultas entre estas variables y cuál es el peso relativo de cada una. Para ello necesitaremos también ciertas nociones de estadística.



Los datos son los siguientes:

ESPECIES	LONGITUD (cm)	ENVERGADURA(cm)	PESO (g)
Ruiseñor común	17.75	25	25
Jilguero europeo	12.5	21.5	19
Pardillo común	12.75	24	19
Verdecillo	11.75	22	14.5
Petirrojo europeo	13.75	21	19
Carbonero común	11.25	18.5	9.5
Herrerillo común	11.5	17.5	9.5
Pinzón vulgar	15	25.5	30
Curruca cabecinegra	11.75	21	12
Mosquitero común	9.75	15	6
Bisbita común	14.5	23	21.5
Lavandera blanca	17.25	28	24.5
Gorrión común	15	23.5	31.5
Canario	12	22	17
Verderón común	12.25	21.5	17
Alondra común	17.25	34.5	65
Currucas capirotadas	13.5	22.5	14
Mosquitero ibérico	10.25	15	6.5
Ruiseñor bastardo	17.75	26.5	27.5
Tarabilla común	12.75	19	15
Agateador europeo	12.75	23	8.5
Carricero común	12.5	20	12
Carricerín común	10.25	18	7.5
Collalba gris	14	22	16.5
Currucas mosquiteras	13.25	21.5	14.5

3.1. Conceptos de estadística:

Supongamos un conjunto de datos de tres variables, por ejemplo longitud (x), envergadura (y) y peso (z) de un conjunto de aves, en este caso 25 especies de aves cantoras. Para cada una de las variables podemos calcular su media y su varianza,

$$\mu_x = \frac{1}{n} \sum x_i \qquad \sigma_x^2 = \frac{1}{n} \sum_i (x_i - \mu_x)^2$$

Para cada una de las variables, es decir, μ_x , μ_y , μ_z , σ_x^2 , σ_y^2 , σ_z^2 . Y definimos una variable adicional, **la covarianza** o varianza conjunta, definida como,

$$\sigma_{xy} = \frac{1}{n} \sum_i (x_i - \mu_x) \cdot (y_i - \mu_y)$$

Nos permite averiguar si hay alguna relación estadística entre dos variables. Como tenemos tres variables las covarianzas serán,

$$\sigma_{xy}, \sigma_{xz}, \sigma_{yz}$$

ya que son magnitudes simétricas. Usando varianzas y covarianzas construimos **la matriz de covarianza**, viene dada por:

$$M = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_z^2 \end{pmatrix}$$

Realizaremos un Análisis de Componentes Principales.

3.2. Cuestiones.

- Introduce los datos en variables de MATLAB.
- Calcula la matriz de covarianza.
- Obtén los valores y vectores propios de dicha matriz.
- Interpreta los resultados obtenidos sobre la base del Análisis de Componentes Principales.
- Escribe un párrafo con tus conclusiones expresando tu punto de vista sobre los resultados.

Problema 4

Estadística de hojas de árbol

La gran diversidad de formas y tamaños de las hojas de los árboles supone un reto. Los parámetros que se utilizan para caracterizar una hoja de una planta pueden variar según la especie de planta y el propósito de la investigación, pero aquí hay algunos parámetros comunes que se utilizan para describir la estructura y las propiedades de las hojas, serían los siguientes:

- Área foliar
- Longitud y ancho de la hoja
- Espesor de la hoja
- Peso de la hoja
- Concentración de clorofila.
- Área específica de la hoja (área foliar por unidad de peso seco)
- Densidad de la hoja
- Índice de área foliar (IAF): es una medida de la cantidad total de hojas en una planta y se calcula dividiendo el área foliar total por el área del suelo ocupada por la planta.



En este problema vamos a trabajar con tres de estas variables, en concreto la longitud, la anchura y la concentración de clorofila de un conjunto de especies. Los datos obtenidos son los siguientes:

VARIABLES

ESPECIES		LONGITUD	ANCHURA	CLOROFILA
	Arce	12.5	9.5	0.8
	Roble	10	12.5	0.5
	Olmo	8	12.5	0.9
	Abedul	6	7.5	0.3
	Sauce	6	10	0.8
	Haya	10	14	0.9
	Chopo	8	17.5	0.3
	Nogal	7.5	17.5	0.9
	Encina	4.5	7.5	0.9
	Olivo	6	9	0.9
	Ciprés	4	10	0.2
	Granado	6	10	0.6
	Manzano	6	10	0.8
	Melojo	10	12.5	0.5
	Rebollo	8	15	0.4
	Castaño	10	15	0.7
	Plátano	12.5	17.5	0.5

4.1. Conceptos de estadística:

Supongamos un conjunto de datos de tres variables, por ejemplo longitud (x), envergadura (y) y peso (z) de un conjunto de aves, en este caso 25 especies de aves cantoras. Para cada una de las variables podemos calcular su media y su varianza,

$$\mu_x = \frac{1}{n} \sum x_i \qquad \sigma_x^2 = \frac{1}{n} \sum_i (x_i - \mu_x)^2$$

Para cada una de las variables, es decir, μ_x , μ_y , μ_z , σ_x^2 , σ_y^2 , σ_z^2 . Y definimos una variable adicional, **la covarianza** o varianza conjunta, definida como,

$$\sigma_{xy} = \frac{1}{n} \sum_i (x_i - \mu_x) \cdot (y_i - \mu_y)$$

Nos permite averiguar si hay alguna relación estadística entre dos variables. Como tenemos tres variables las covarianzas serán,

$$\sigma_{xy}, \sigma_{xz}, \sigma_{yz}$$

ya que son magnitudes simétricas. Usando varianzas y covarianzas construimos **la matriz de covarianza**, viene dada por:

$$Cov = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_z^2 \end{pmatrix}$$

Realizaremos un Análisis de Componentes Principales.

4.2. Cuestiones.

- Introduce los datos en variables de MATLAB.
- Calcula la matriz de covarianza.
- Obtén los valores y vectores propios de dicha matriz.
- Interpreta los resultados obtenidos sobre la base del Análisis de Componentes Principales.
- Escribe un párrafo con tus conclusiones expresando tu punto de vista sobre los resultados.

Problema 5

Propagación epidémica.

(6.0 puntos) Un modelo de propagación de epidemias distingue tres tipos de individuos, los denominaremos **sanos (x)**, **infectados (y)** y **recuperados (z)**. Se ha establecido que la aplicación lineal que relaciona la evolución de la población de sanos, infectados y recuperados entre un instante de tiempo y el siguiente es la que sigue:

$$f(x, y, z) = ((1 - m)x, mx + (1 - c)y, cy + z)$$

Donde m y c son parámetros que inicialmente tomaremos como $m = 0,01$ y $c = 0,02$, se pide lo siguiente:

- (1.0 puntos) Calcula la matriz de transición entre una generación y la siguiente.
- (1.5 puntos) Si la población inicial viene dada por $\vec{u} = (\text{sanos}, \text{infectados}, \text{recuperados}) = (9000, 1000, 0)$ calcula la distribución de individuos en la segunda, décima y centésima generación.
- (1.5 puntos) Calcula mediante diagonalización la población en la centésima generación y en la milésima generación.
- (1.0 puntos) Si consideramos que cada intervalo de tiempo corresponde con 1 día. ¿En que día la población de recuperados supera a los sanos?
- (0.5 puntos) ¿Y en que día la población de infectados supera a los sanos? ¿por qué?
- (0.5 puntos) Siendo $m = 0,01$, ¿para qué valor del parámetro c los recuperados superan a los infectados el día 29?

Problema 6

Estadística de notas.

(4.0 puntos) Un equipo de estadísticos ha tomado datos de los resultados de EBAU para 3 asignaturas y una muestra aleatoria de 100 alumnos. Se quiere completar un Análisis de Componentes Principales con estos datos. Se pide,

- (1.0 puntos) Introduce los datos en Matlab y calcula las medias.
- (1.0 puntos) Calcula la matriz de covarianza y diagonaliza.
- (1.0 puntos) Interpreta los resultados obtenidos.
- (1.0 puntos) ¿Qué porcentaje de la varianza explica la primera componente principal?

Los datos para el análisis se adjuntan en un archivo excel y son los siguientes

Alumnos	Asig1	Asig2	Asig3	Alumnos	Asig1	Asig2	Asig3
1	5.973	6.7739	6.5826	47	6.4131	5.7014	6.6428
2	7.2972	5.2483	5.1104	48	5.2969	4.5935	5.9973
3	5.332	5.3983	6.0564	49	5.4317	3.3078	6.0605
4	5.2323	3.7882	5.8465	50	4.7652	1.8356	6.4868
5	7.7061	5.104	6.4708	51	3.2032	4.8397	5.1964
6	4.5792	4.5847	6.4914	52	5.238	6.5798	5.6842
7	5.296	4.7786	6.4867	53	6.6588	4.927	7.0113
8	6.6009	3.3784	5.9428	54	7.7582	4.1167	6.6228
9	5.5119	2.3288	5.7763	55	4.6916	5.2628	5.294
10	7.2869	3.7488	5.7831	56	2.9986	2.2235	4.6421
11	6.7625	3.0973	6.8035	57	5.5353	3.6422	5.7961
12	5.7123	4.3791	4.8897	58	5.8321	5.4516	6.3513
13	4.2002	3.0464	6.0534	59	4.979	2.243	6.8664
14	6.1014	4.05	6.3912	60	3.6752	0.89107	6.2663
15	3.3447	1.4555	5.7384	61	8.1628	6.0466	5.4064
16	5.6457	5.2276	6.5375	62	1.9631	2.9224	6.4165
17	6.4433	5.2098	6.1496	63	4.5788	3.8559	5.371
18	5.6013	4.1409	5.5365	64	3.4309	3.7738	5.5205
19	6.5441	5.9727	6.3141	65	4.4004	4.9757	5.7377
20	8.1834	4.4713	6.0104	66	5.7495	5.5133	5.3307
21	6.6066	5.2612	5.9102	67	5.4952	3.5013	6.0084
22	5.6775	4.7155	5.697	68	4.2949	4.5894	5.5636
23	6.2074	4.9815	6.2434	69	6.3988	4.8616	5.998
24	5.9017	3.5077	4.8946	70	6.1802	3.9209	5.873
25	3.4407	3.5706	5.9183	71	6.2155	3.0963	5.9102
26	4.7111	3.175	5.8157	72	4.6981	4.5914	5.8439
27	7.032	3.8681	5.5552	73	6.7829	4.4084	7.0721
28	5.2384	2.5955	6.6615	74	6.2344	4.6241	6.6316
29	6.1237	5.7829	5.6808	75	8.8479	4.9028	6.2418
30	3.8766	2.3522	6.5823	76	6.804	2.953	5.1627
31	5.3912	3.9338	6.89	77	6.9361	5.1282	5.0504
32	6.1489	3.0442	6.0891	78	5.3749	4.0614	4.8107
33	3.3745	2.7734	6.3964	79	3.4779	4.6575	8.0422
34	5.7916	0.38736	6.6522	80	4.775	4.8194	6.8876
35	5.0428	2.6402	6.1427	81	4.2283	4.8542	6.3817
36	3.9713	1.6886	6.5825	82	4.5024	2.784	5.5111
37	5.2469	2.831	5.3072	83	6.0851	3.3458	5.3894
38	2.6182	2.4979	6.9852	84	4.5126	2.152	6.5872
39	2.2906	2.1968	5.5003	85	4.0341	4.077	7.3067
40	3.6356	5.6492	4.9068	86	4.7211	7.2265	5.0026
41	3.8992	2.6993	7.179	87	5.0725	4.8773	5.8822
42	3.6411	2.9643	6.1556	88	5.8789	3.4426	6.1404
43	2.8599	0.95793	6.5174	89	3.9528	4.451	4.7914
44	5.7076	3.7191	5.9431	90	5.7488	2.9425	6.7949
45	3.0455	1.3341	5.4783	91	5.4103	2.4106	5.9068
46	5.4559	4.215	4.0271	92	7.173	5.9152	7.5521
...							

Alumnos	Asignatura1	Asignatura2	Asignatura3
93	4.9985	2.9965	5.9236
94	5.1098	4.3919	6.4893
95	3.6435	4.9944	5.3922
96	3.3409	3.6117	5.7582
97	5.7391	5.4582	5.3736
98	4.9472	1.9569	6.4776
99	6.0124	5.2664	7.1205
100	4.0376	6.0879	5.7802



Este obra está bajo una [licencia de Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/).