

Javier García Tercero y Lucian Andrei Negoita

Tarea 2.

Ejercicios,

tema 3

CASOS: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 y 13
4 de marzo, 2025

Curso 2024-2025. UCLM, Albacete
Integración de Sistemas Informáticos
3º Curso del Grado en Ingeniería Informática
Grupo A

Índice

Problema 1.....	3
Problema 2.....	5
Problema 3.....	6
Problema 4.....	8
Problema 5.....	9
Problema 6.....	10
Problema 7.....	11
Problema 8.....	13
Problema 9.....	15
Problema 10.....	16
Problema 11.....	18
Problema 12.....	19
Problema 13.....	20

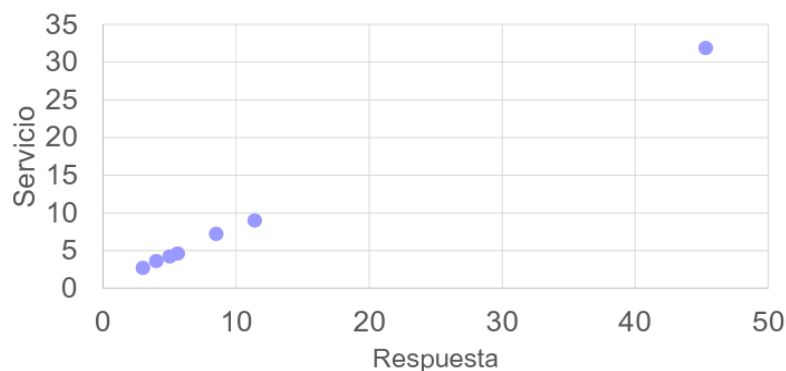
Problema 1

(Resuelto en clase)

- En la tabla siguiente se indican las mediciones del tiempo de respuesta (t_r) y del tiempo de servicio (t_s) realizadas en más de doscientos subsistemas de disco con distintas cantidades de memoria cache. Se pide agrupar estos datos en tres clusters usando distancias euclídeas y dibujar el dendograma que se produzca

Clase	Tamaño de cache (GB)	t_r (ms)	t_s (ms)
1	0	45.3	31.9
2	1	11.4	9
3	2	8.5	7.2
4	3	5.6	4.6
5	4	5	4.2
6	8	4	3.6
7	16	3	2.7

- Gráficamente



- Partimos de 7 clases y calculamos la matriz de distancias

$$d_{C1C2} = \sqrt{(45.3 - 11.4)^2 + (31.9 - 9)^2} = 40.9$$

$$d_{C4C6} = \sqrt{(5.6 - 4)^2 + (4.6 - 3.6)^2} = 1.88$$

Clase	C1	C2	C3	C4	C5	C6	C7
C1	0	40.9	44.3	48.2	48.9	50.1	51.4
C2		0	3.4	7.3	8.0	9.2	10.5
C3			0	3.9	4.6	5.8	7.1
C4				0	0.7	1.9	3.2
C5					0	1.2	2.5
C6						0	1.3
C7							0

- Agrupamos C4 y C5 y calculamos el centroide de la nueva clase C45: $\left(\frac{5.6 + 5}{2}, \frac{4.6 + 4.2}{2}\right) = (5.3, 4.4)$

- Calculamos la nueva matriz de distancias:

Clase	C1	C2	C3	C45	C6	C7
C1	0	40.9	44.3	48.5	50.1	51.4
C2		0	3.4	7.6	9.2	10.5
C3			0	4.3	5.8	7.1
C45				0	1.5	2.9
C6					0	1.3
C7						0

- Agrupamos C6 y C7 y calculamos el centroide de la nueva clase C67: $\left(\frac{4 + 3}{2}, \frac{3.6 + 2.7}{2}\right) = (3.5, 3.2)$

- Calculamos la nueva matriz de distancias:

Clase	C1	C2	C3	C45	C67
C1	0	40.9	44.3	48.5	50.7
C2		0	3.4	7.6	9.8
C3			0	4.3	6.4
C45				0	2.2
C67					0

- Agrupamos C45 y C67 y calculamos el centroide de la nueva clase C4567: $\left(\frac{5.3 + 3.5}{2}, \frac{4.4 + 3.2}{2}\right) = (4.4, 3.8)$

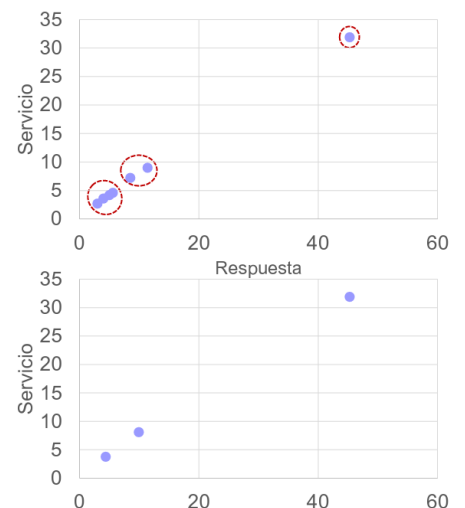
- Calculamos la nueva matriz de distancias:

Clase	C1	C2	C3	C4567
C1	0	40.9	44.3	49.6
C2		0	3.4	8.7
C3			0	5.3
C4567				0

- Agrupamos C2 y C3 y calculamos el centroide de la nueva clase C23: $\left(\frac{11.4 + 8.5}{2}, \frac{9 + 7.2}{2}\right) = (10, 8.1)$

- Calculamos la nueva matriz de distancias:

Clase	C1	C23	C4567
C1	0	42.6	49.6
C23		0	7.0
C4567			0



Problema 2

Problema 2

Se ha monitorizado un servidor web a lo largo de diez meses. Durante el periodo de monitorización se ha grabado en un fichero de bitácora la utilización del procesador, la del disco y el tiempo de ejecución en modo usuario. En la tabla siguiente se muestran los valores medios de esos tres parámetros obtenidos tras procesar el fichero de monitorización y agrupar las solicitudes mediante un criterio cualitativo.

Solicitud	Procesador (%)	Disco (%)	Ejecución (ms)
1	8.4	30	700
2	15.0	10	140
3	15.3	11	150
4	8.5	50	60
5	11	90	115
6	16	50	260
7	26	50.3	4000
8	50	50	150
9	80	50	100
10	50	30	150

Hágase una caracterización de la carga utilizando la distancia euclídea como métrica de agrupamiento. Téngase en cuenta que el tiempo de ejecución es directamente proporcional al tamaño del documento.

Tabla

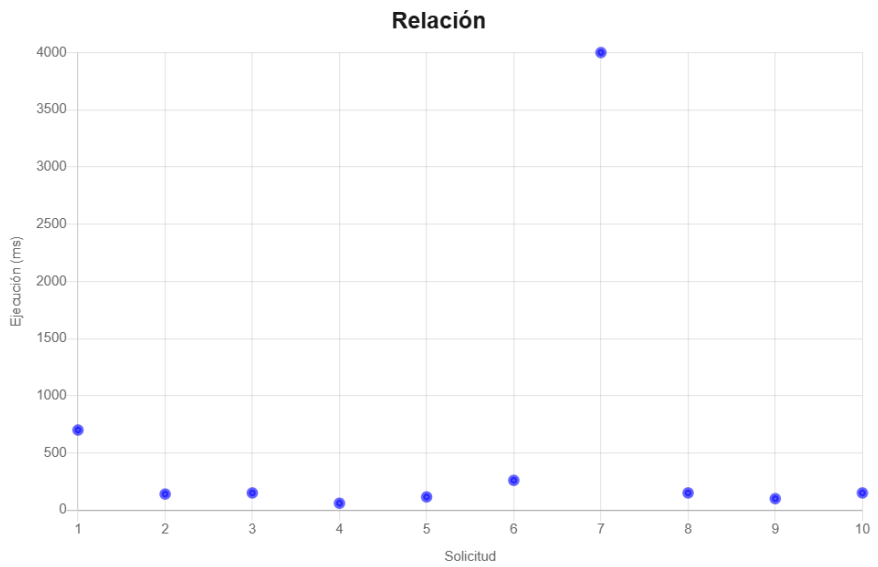
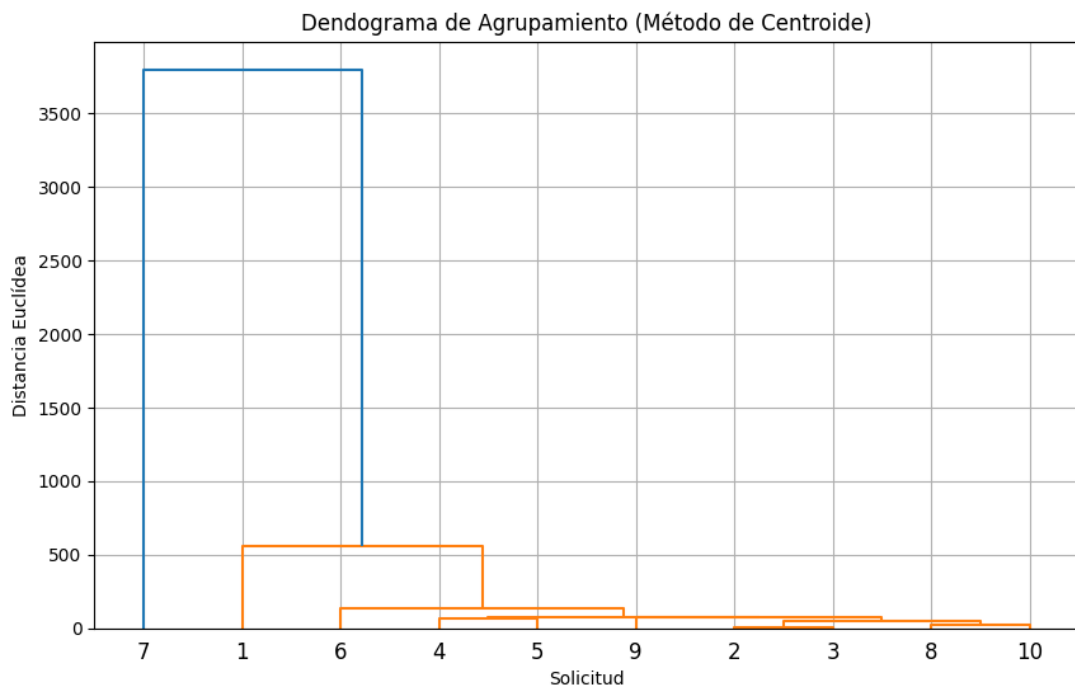


Tabla de distancias

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
S1	0	560,35	550,32	640,31	588,06	440,45	3300,06	550,36	600,33	550
S2		0	10,04	89,44	83,81	126,49	3860,21	41,23	56,56	22,36
S3			0	98,08	86,40	116,70	3850,20	39	63,41	19
S4				0	68,00	200	3940,00	90	40	92,19
S5					0	150,41	3885,20	53,15	42,72	96,56
S6						0	3740,00	110	160	111,80
S7							0	3850,00	3900,00	3850,05
S8								0	50	20
S9									0	53,85
S10										0

Dendograma

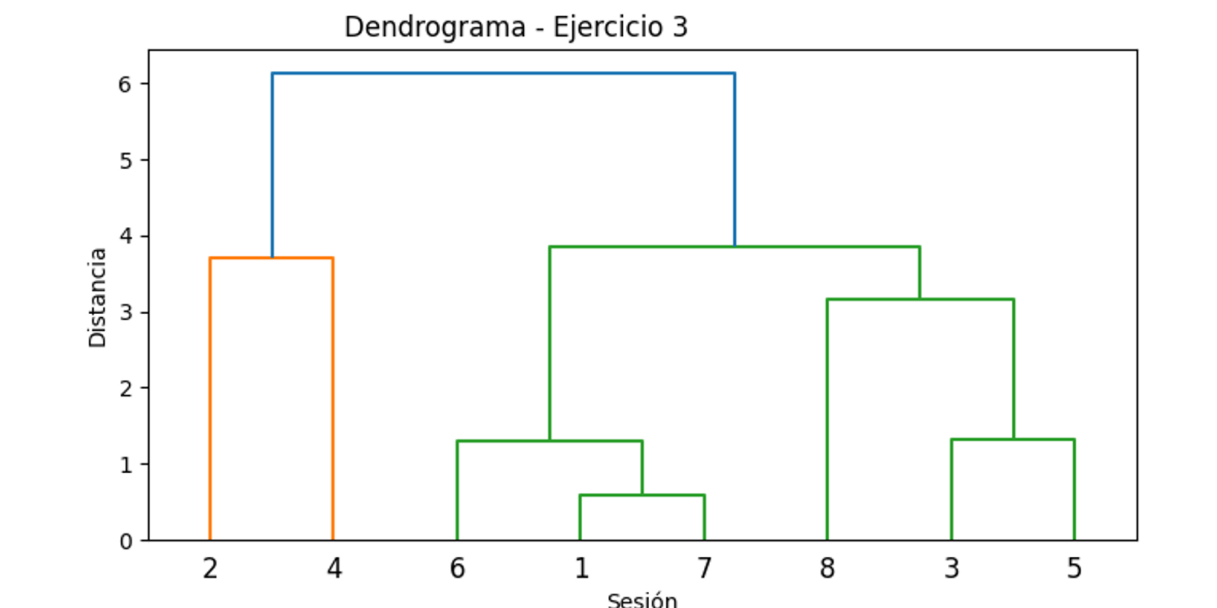


Problema 3

En la tabla siguiente se presenta un pequeño conjunto de muestras de la tasa de visitas correspondiente a ocho sesiones de acceso a un servidor web de comercio electrónico por parte de un mismo usuario. Agrupe esta carga hasta un nivel que parezca razonable.

Sesión	V_{Muestra}	$V_{\text{Búsqueda}}$	$V_{\text{Añade al carro}}$	$V_{\text{Selecciona}}$	V_{Paga}
1	15	12	2	5	1
2	10	15	1	14	0
3	4	17	2	4	1
4	18	20	4	15	0
5	4	12	2	7	1
6	16	11	3	7	1
7	17	12	2	7	1
8	5	4	1	2	1

Llevando a cabo el agrupamiento obtenemos este dendrograma, del agrupamiento es mediante clustering.



Problema 4

Problema 4

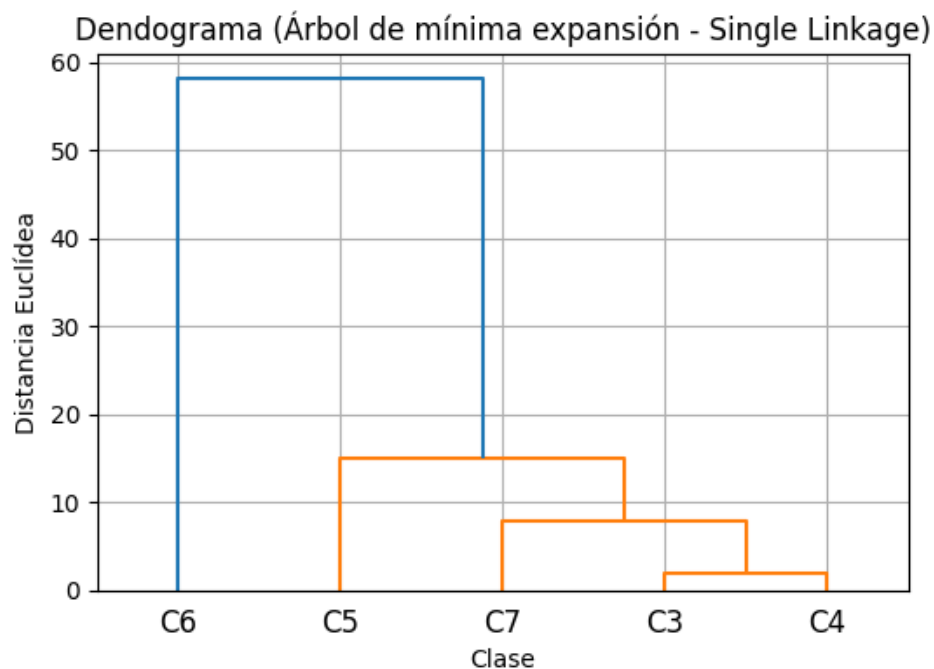
El tiempo de procesador y el número de operaciones de E/S de cinco de las siete clases de programas que trabajan contra una base de datos relacional y unos programas de edición de textos se muestran en la tabla siguiente. Se pide preparar un dendograma utilizando un algoritmo de árbol de mínima expansión para el análisis de agrupamiento.

Clase	Programa	Procesador (ms)	E/S por s
C3	Actualizaciones pesadas	8	27
C4	Actualizaciones medianas	6	27
C5	Actualizaciones ligeras	6	12
C6	Consulta pesada	4	91
C7	Editor de textos ligero	1	33

Tabla

	C3	C4	C5	C6	C7
C3	0	2	15,13	64,12	9,22
C4		0	15	64,03	7,81
C5			0	79,03	21,59
C6				0	58,08
C7					0

Dendograma



(En el árbol, la altura a la que se fusionan dos nodos corresponde a la distancia entre sus centroides)

Problema 5

Problema 5

Se han realizado una serie de mediciones durante los días laborables de una semana en la hora punta del proceso de una transacción tipo que se ejecuta en un computador de la central de reservas aéreas y se han detectado los consumos medios de los recursos según se presentan en la siguiente tabla. Se pide realizar un agrupamiento hasta llegar a dos clases representativas de la carga. Indica los centroides de las clases resultantes.

Solicitud	Procesador (ms)	E/S (ms)	Red (ms)
Lunes	18	25	9
Martes	17	28	8
Miércoles	20	30	10
Jueves	19	29	9
Viernes	22	33	8

Tabla reordenada

	C1	C2	C3	C4	C5
Procesador	18	17	20	19	22
E/S	25	28	30	29	33
Red	9	8	10	9	8

Tabla de distancias inicial

	C1	C2	C3	C4	C5
C1	0	3,32	5,48	4,12	9
C2		0	4,12	2,45	7,07
C3			0	1,73	4,12
C4				0	5,1
C5					0

La amarilla es la distancia mínima, se une c3 y c4

	C1	C2	C34	C5
C1	0	3,32	4,77	9
C2		0	3,28	7,07
C34			0	4,56
C5				0

La amarilla es la distancia mínima, se une c34 y c2

	C1	C234	C5
C1	0	3,77	9
C234		0	5,72
C5			0

La amarilla es la distancia mínima, se une c234 y c1

Columna1	C1234	C5
C1234	0	7,3
C5		0

Ya no hay más reducción por lo que pide el ejercicio, ya que el resultado debe dar 2 columnas/clases representativas

Centroide de C1234 = (18,125, 26,875, 8,875)

Centroide de C5 = (22,33,8)

Problema 6

Un servidor de aplicaciones ha sido monitorizado durante una semana. Durante ese periodo se grabó un fichero bitácora que contiene, entre otros, el número de lecturas y escrituras en el disco duro (operaciones) así como el porcentaje de utilización del procesador. En la tabla siguiente se recogen los valores monitorizados esa semana. Hágase una agrupación de los siete días en uno que caracterice ambos parámetros.

Día	Uso del procesador (%)	Operaciones de E/S
1	30	70
2	50	25
3	50	250
4	70	150
5	12	15
6	40	25
7	50	20

Como es un único perfil usaremos clustering con $k=1$.

Día	Uso Procesador	Operaciones E/S
Lunes	30	70
Martes	50	25
Miércoles	50	250
Jueves	70	150
Viernes	12	15
Sábado	40	25
Domingo	50	20

Perfil representativo:

Uso_Procesador 43.142.857

Operaciones_E/S 79.285.714

Problema 7

Problema 7

En la tabla siguiente se muestra el número de operaciones de E/S por segundo y el tamaño del fichero solicitado por cinco clases de transacciones web que resumen la carga de un servidor de aplicaciones:

Clase	Nombre de la transacción	Operaciones/s	Tamaño (MB)
C1	Ficheros muy grandes	9	9
C2	Ficheros grandes	6	1
C3	Ficheros medianos	5	0.25
C4	Ficheros pequeños	4	0.02
C5	Ficheros muy pequeños	3	0.001

Prepárese un dendograma utilizando un algoritmo de árbol de extensión mínima para el análisis de agrupamiento. Determinése el grado de agrupamiento óptimo en un número de clases menor.

Muy similar al 5

Tabla completa

	C1	C2	C3	C4	C5
C1	0	8,54	9,62	10,28	10,82
C2		0	1,25	2,23	3,16
C3			0	1,03	2,02
C4				0	1
C5					0

Tabla unión C4 y C5

	C1	C2	C3	C45
C1	0	8,54	9,62	10,54
C2		0	1,25	2,69
C3			0	1,52
C45				0

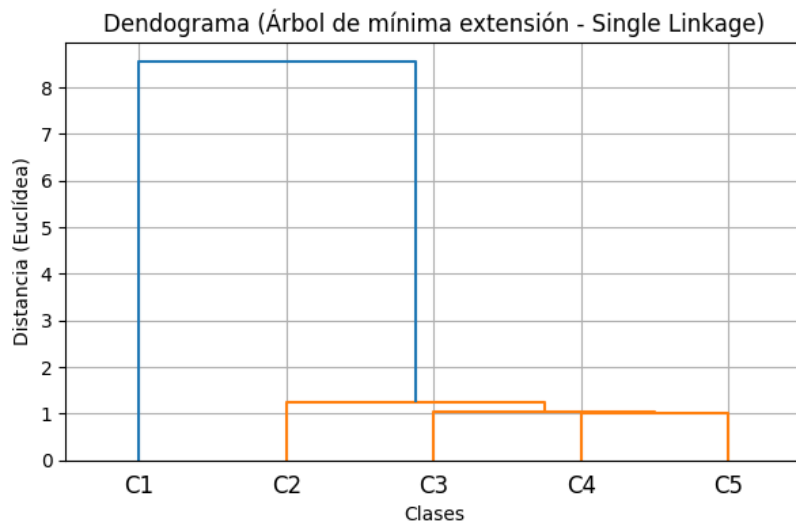
Tabla unión C2 y C3

	C1	C23	C45
C1	0	9,08	10,54
C23		0	2,09
C45			0

Tabla unión C23 C45

Columna1	C1	C2345
C1	0	9,78
C23		0

Dendograma



Problema 8

En un sistema informático se procesa una carga real W , generándose a partir de ella dos modelos, W_1 y W_2 . Hay tres tipos de aplicaciones, una de cálculo (C), otra de lectura (L) y una tercera intermedia (I), según se especifica en la tabla siguiente. En ella se indica el tiempo de procesador, el tiempo de presencia y el número de operaciones de E/S lógicas. Los tiempos vienen expresados en segundos.

Parámetros	WC	WL	WI	W_1C	W_1L	W_1I	W_2C	W_2L	W_2I
Procesador	80	30	40	90	25	42	190	15	80
Presencia	1000	400	500	1100	350	530	2000	500	250
Operaciones E/S	300	5000	1000	600	2500	500	280	4000	900

Los rangos asignados a cada parámetro de la carga se indican a continuación:

Parámetros	Min	Max	Rango
Procesador	20	220	200
Presencia	200	2200	2000
Operaciones E/S	100	3100	3000

Dadas las características de ambos modelos, se pide analizar cuál de los dos explica mejor la carga real W, de acuerdo con los objetivos y criterios de ponderación que se exponen a continuación:

1. Objetivo 1. Dar más importancia a la actividad del procesador y al tiempo de presencia ($w_{11}=2$, $w_{21}=2$, $w_{31}=0$).

2. Objetivo 2. Dar más importancia a la actividad de E/S ($w_{12}=0$, $w_{22}=1$, $w_{32}=2$).

La variable w_{ij} representa el peso del parámetro i en el objetivo j . Los porcentajes de programas existentes de cada clase son: $q_C=40\%$, $q_L=40\%$ y $q_I=20\%$.

Para cada parámetro vamos a normalizar con $v_{norm} = v - v_{min} / v_{max} - v_{min}$, donde v_{min} y v_{max} son los rangos dados.

Sacamos la diferencia en cada valor (normalizado frente a real). Y finalmente multiplicamos cada diferencia por peso y sumaremos las diferencias, cogiendo al de menor distancia como mejor.

	Carga Real	W1	W2
Procesador	80	90	190
Presencia	1000	1100	2000
Operaciones E/S	300	600	280

Datos normalizados

	Carga Real	W1	W2
Procesador	0,3	0,35	0,85
Presencia	0,4	0,45	0,9
Operaciones E/S	0,067	0,167	0,06

Distancias ponderadas Objeto 1

Columna1	Carga Real	W1	W2
W1:	0,2		
W2:	2		

Problema 9

La tabla siguiente contiene los valores de los parámetros que caracterizan la carga real W de un sistema informático y dos modelos W_1 y W_2 de representación de la misma.

Parámetro	$WC1$	$WC2$	$WC3$
% de la carga total	50%	30%	20%
Tiempo de procesador (s)	100	50	75
Operaciones de E/S	200	1000	700
Memoria principal (MB)	300	550	400
Ficheros de disco	10	6	7
Páginas impresas	2000	3000	2300
Parámetro	W_1C1	W_1C2	W_1C3
Tiempo de procesador (s)	100	50	80
Operaciones de E/S	240	1350	950
Memoria principal (MB)	300	560	410
Ficheros de disco	11	10	12
Páginas impresas	2050	3100	2200
Parámetro	W_2C1	W_2C2	W_2C3
Tiempo de procesador (s)	140	70	95
Operaciones de E/S	200	1000	800
Memoria principal (MB)	330	580	410
Ficheros de disco	13	8	10
Páginas impresas	2100	3200	2400

Los rangos se indican en la tabla siguiente:

Parámetro	Max	Min
Tiempo de procesador (s)	200	0
Operaciones de E/S	2100	100
Memoria principal (MB)	1000	100
Ficheros de disco	20	0
Páginas impresas	4100	100

Los rangos se indican en la tabla siguiente:

Parámetro	Max	Min
Tiempo de procesador (s)	200	0
Operaciones de E/S	2100	100
Memoria principal (MB)	1000	100
Ficheros de disco	20	0
Páginas impresas	4100	100

Se pide analizar cuál de los dos modelos caracteriza mejor la carga real para cada uno de los siguientes objetivos:

- Objetivo 1. Con respecto a la actividad del procesador, dando los siguientes pesos a cada uno de los cinco parámetros: $w_1=2$, $w_2=1$, $w_3=1$, $w_4=1$ y $w_5=0.5$.
- Objetivo 2. Con respecto a la actividad de operaciones de E/S, dando los siguientes pesos a cada uno de los cinco parámetros: $w_1=1$, $w_2=2$, $w_3=1$, $w_4=2$ y $w_5=0.5$.
- Objetivo 3. Con respecto a la actividad conjunta de E/S, procesador y memoria principal, dando los siguientes pesos a cada uno de los cinco parámetros: $w_1=1$, $w_2=1$, $w_3=1$, $w_4=1$ y $w_5=1$.

Normalización y distancias con la calculadora
Aquel con menor carga será más representativo.

Objetivo 1 (CPU)

Distancia W1 = 0.2474

Distancia W2 = 0.4901

→ El modelo W1 representa mejor la carga real para este objetivo.

Objetivo 2 (E/S)

Distancia W1 = 0.4649

Distancia W2 = 0.4851

→ El modelo W1 representa mejor la carga real para este objetivo.

Objetivo 3 (Conjunto)

Distancia W1 = 0.2518

Distancia W2 = 0.3564

→ El modelo W1 representa mejor la carga real para este objetivo.

Problema 10

La tabla siguiente contiene los valores de los parámetros que caracterizan la carga real W de un servidor y dos modelos W1 y W2 de representación de la misma.

Parámetro	WC1	WC2
% de la carga total	50%	50%
Tiempo de procesador (s)	100	20
Tiempo de E/S (s)	20	100
Tiempo de red (s)	10	80
Parámetro	W ₁ C1	W ₁ C2
Tiempo de procesador (s)	95	18
Tiempo de E/S (s)	15	90
Tiempo de red (s)	8	75
Parámetro	W ₂ C1	W ₂ C2
Tiempo de procesador (s)	90	15
Tiempo de E/S (s)	18	95
Tiempo de red (s)	7	70

Los rangos se indican en la tabla siguiente:

Parámetro	Max	Min
Tiempo de procesador (s)	100	5
Tiempo de E/S (s)	100	5
Tiempo de red (s)	90	5

Se pide analizar cuál de los dos modelos caracteriza mejor la carga real para cada uno de los siguientes objetivos:

- Objetivo 1. Con respecto a la actividad del procesador, dando los siguientes pesos a cada uno de los tres parámetros: $w_1=2$, $w_2=0.5$ y $w_3=0.5$.
- Objetivo 2. Con respecto a la actividad de E/S, dando los siguientes pesos a cada uno de los tres parámetros: $w_1=0.5$, $w_2=2$ y $w_3=0.5$.
- Objetivo 3. Con respecto a la actividad conjunta de procesador, E/S y red, dando los siguientes pesos a cada uno de los tres parámetros: $w_1=1$, $w_2=1$ y $w_3=1$.

Normalizamos los datos con la fórmula de siempre $v\text{-norm} = (v - \min) / (\max - \min)$

Calculamos las diferencias entre W1 y W2 y la carga real.
Se multiplican los pesos y se suman las diferencias ponderadas y cogemos la menor.

Datos normalizados

Columna1	Carga real	W1	W2
Procesador	1	0,95	0,9
E/S	0,16	0,11	0,14
Red	0,06	0,04	0,23

Distancias ponderadas, objetivo 1 - Procesador

W1 = 0,143

W2 = 0,239

Distancias ponderadas (Objetivo 2 - E/S)

W1 = 0,143

W2 = 0,112

Distancias ponderadas (Objetivo 3 -Actividad conjunta)

W1 = 0,129

W2 = 0,162

Problema 11

La tabla siguiente contiene los valores de los parámetros que caracterizan la carga real de un sistema informático W (constituida por tres clases de componentes, WC1, WC2 y WC3) y dos modelos de representación de la misma, W1 y W2.

Parámetro	WC1	WC2	WC3
% de la carga total	27%	31%	36%
Tiempo de CPU (s)	80	60	5
Operaciones de E/S de disco	150	1200	40
Memoria principal (KBytes)	40	90	20
Ficheros de disco	2.5	4.1	1.8
Líneas impresas	300	1100	50
Parámetro	W ₁ C1	W ₁ C2	W ₁ C3
Tiempo de CPU (s)	75	62	4
Operaciones de E/S de disco	144	1220	51
Memoria principal (KBytes)	38	89	18
Ficheros de disco	2.2	4.2	2
Líneas impresas	220	1250	65
Parámetro	W ₂ C1	W ₂ C2	W ₂ C3
Tiempo de CPU (s)	71	64	9
Operaciones de E/S de disco	146	1190	48
Memoria principal (KBytes)	40	86	18
Ficheros de disco	2.4	4.1	2.1
Líneas impresas	320	1180	48

Los rangos se indican en la tabla siguiente:

Parámetro	Max	Min
Tiempo de CPU (s)	190	0
Operaciones de E/S de disco	1650	30
Memoria principal (KBytes)	180	8
Ficheros de disco	6	0
Líneas impresas	2250	6

Se pide analizar cuál de los dos modelos caracteriza mejor la carga real para cada uno de los siguientes objetivos:

- Objetivo 1. Con respecto a la actividad de E/S de disco, dando los siguientes pesos a cada uno de los cinco parámetros: $w_1=1$, $w_2=2$, $w_3=1$, $w_4=2$ y $w_5=0.5$.

- Objetivo 2. Con respecto a la actividad de la CPU, dando los siguientes pesos a cada uno de los cinco parámetros: $w_1=2$, $w_2=1$, $w_3=1$, $w_4=1$ y $w_5=0.5$.

Usamos la fórmula de siempre otra vez para normalizar, luego sacamos diferencia entre lo normalizado y la carga real. Sumamos diferencias para distancia total en cada modelo y luego comparamos los modelos, siendo el de menor distancia el que representa la carga real para ese objetivo. Recordando siempre que la diferencia es en valor absoluto.

Objetivo 1 (Actividad de E/S de disco):

Modelo W1: 0.114

Modelo W2: 0.099

Objetivo 2 (Actividad de CPU):

Modelo W1: 0.088

Modelo W2: 0.099

Para Objetivo 1, el modelo que mejor representa la carga real es: W2

Para Objetivo 2, el modelo que mejor representa la carga real es: W1

Problema 12

Se ha realizado una serie de mediciones, durante los días laborables de una semana, en el cajero de un banco y se ha anotado el consumo de ciertos recursos según se muestra en la tabla. Se pide hacer un agrupamiento hasta llegar a una sola clase que represente un día laborable cualquiera.

Solicitud	Modo usuario (min)	Modo espera (s)	Modo red (s)
Lunes	19	25	6
Martes	17	28	8
Miércoles	20	30	10
Jueves	18	29	9
Viernes	22	31	12

Sacamos la media de cada parámetro y así conseguimos la clase, porque obtenemos el centroide.

	Modo_usuario	Modo_espera	Modo_red
Lunes	19	25	6
Martes	17	28	8
Miércoles	17	30	10
Jueves	18	29	9
Viernes	22	31	12

Perfil representativo del cajero:	
Modo usuario	19.2
Modo espera	28.6
Modo red	9.0

Problema 13

Problema 13

Consideremos un servidor web dedicado a la entrega de documentos HTML bajo demanda. El servidor se ha monitorizado durante un periodo de tiempo determinado, obteniendo la tabla siguiente, donde se especifica el valor de los dos parámetros que caracterizan la carga: tamaño de los documentos solicitados y número de accesos.

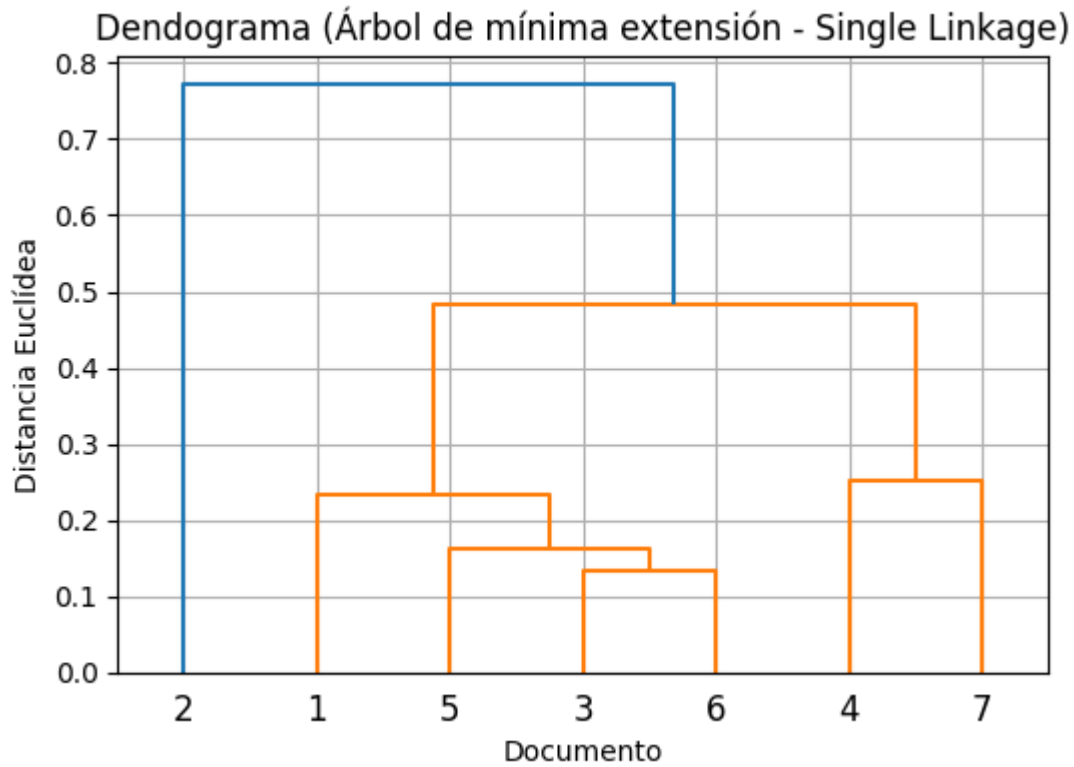
Documento	Tamaño (kb)	Número de accesos
1	1.08	2.45
2	2.18	1.45
3	0.70	2.47
4	1.40	2.09
5	0.85	2.41
6	0.60	2.38
7	1.54	1.88

Se pide hacer un estudio de agrupamiento que obtenga tres clases representativas de la carga en el servidor, suponiendo que las observaciones ya han sido escaladas, y empleando como mecanismo de agrupamiento el árbol de mínima expansión y como métrica de distancia la distancia euclídea. Indíquese el centroide de las clases resultantes.

Tabla con las distancias

	C1	C2	C3	C4	C5	C6	C7
C1	0	1,49	0,38	0,48	0,23	0,49	0,73
C2		0	1,8	1	1,64	1,83	0,77
C3			0	0,8	0,16	0,13	1,03
C4				0	0,64	0,85	0,25
C5					0	0,25	0,87
C6						0	1,06
C7							0

(Omitimos rehacer el simplificado por repetición)



Docu.	Tamaño	Acceso	Cluster
1	1,08	2,45	1
2	2,18	1,45	3
3	0,7	2,47	1
4	1,4	2,09	2
5	0,85	2,41	1
6	0,6	2,38	1
7	1,54	1,88	2

Centroides

Cluster 1 = Tamaño=0.807, Accesos=2.428

Cluster 2 = Tamaño=1.470, Accesos=1.985

Cluster 3 = Tamaño=2.180, Accesos=1.450