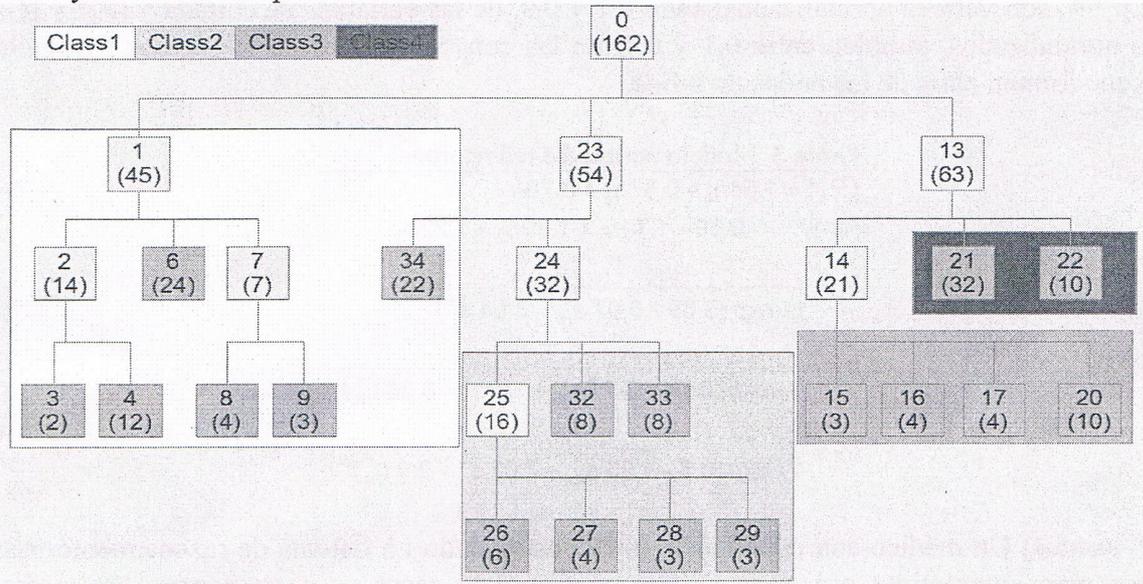


## INTRODUCCIÓN AL APRENDIZAJE AUTOMÁTICO 11 Junio 2014

Alumno/a D.....

**Cuestiones.-** Analice las siguientes cuestiones y justifique en su caso las respuestas

**1) (2 puntos)** Explique cual es la técnica aplicada por el algoritmo COBWEB. Explique las ventajas e inconvenientes de los algoritmos particionales frente a los jerárquicos. ¿Que medidas de calidad del agrupamiento conoce? Analice el dendograma de la figura y justifique el porque considerar 4 clases y cuál es la composición de las mismas.



**2) (1 punto)**

Explique brevemente en que consiste el algoritmo de retropropagación del error en modelos de redes neuronales. Analice el significado de la siguiente regla de actualización de los pesos de la red

$$w_k = w_k + \Delta w_k; \Delta w_k = -\eta \frac{\partial E}{\partial w_k}$$

¿Cómo se obtienen los valores del parámetro  $\eta$ ? ¿Qué significado tiene?

**3) (2 puntos)** Explique como se construye la Tabla 1 donde se muestran las medias y los rangos medios obtenidos para seis bases de datos por nueve algoritmos diferentes. Los resultados del tests de Nemenyi para  $\alpha = 0.10$  y  $\alpha = 0.05$  se pueden ver en la Tabla 2. Explique que hipótesis nulas son las que contrasta el citado test, y analice los datos de la Tabla 2, explicando en función de los resultados cuales son las conclusiones. Cuando se debe de aplicar este test?

**Tabla 1**

Media de	RBFN	MLogistic	SLogistic	C4.5	NaiveBayes	LMT	IB1	SVM	GRBF
RMSE	0,28	0,24	0,30	0,39	0,27	0,30	0,27	0,36	0,18
RangoRMSE	4,47	4,38	4,88	7,58	5,02	5,05	5,05	6,77	1,75

**Tabla 2**

Método (i)	Test de Nemenyi								
	Método (j)								
	RBFN	MLogistic	SLogistic	C4.5	NaiveBayes	LMT	IB1	SVM	GRBF
RBFN	-	0,08	0,41	3,11	0,55	0,58	0,58	2,30	2,72
MLogistic	-	-	0,50	3,19	0,63	0,66	0,66	2,38	2,63
SLogistic	-	-	-	2,69	0,13	0,16	0,16	1,88	3,13
C4,5	-	-	-	-	2,55	2,52	2,52	0,80	5,83
NaiveBayes	-	-	-	-	-	0,02	0,02	1,75	3,27
LMT	-	-	-	-	-	-	0,00	1,72	3,30
IB1	-	-	-	-	-	-	-	1,72	3,30
SVM	-	-	-	-	-	-	-	-	5,02

$CD \alpha=0,10=2,60$  ;  $CD \alpha=0,05=2,83$

### Ejercicios

1.- (2.5 puntos) Dada la siguiente Tabla 3 sitúe los coeficientes del modelo en las conexiones del grafo del modelo de red. ¿Que tipo de modelo de red es? ¿Que estructura tiene? ¿Cuál es el valor asociado al patrón de entrada (2,5, 3,2, 0,7) sabiendo que los mínimos y máximos de las variables  $a_m$ ,  $s$  y  $t_m$  para los patrones de entrenamiento son [1-4] [2-4] [0-1] y de las variables [P] y [Q], [1-4] [2-5]?  $\hat{a}_m^*$ ,  $\hat{s}^*$  y  $\hat{t}_m^*$ , son valores normalizados, entre 0,1 y 0,9, de las variables de entrada y [P]\* y [GA]\* valores normalizados, también entre 0,1 y 0,9, de las concentraciones de los ácidos Pyrogallol y Gálico que forman parte de los nodos de salida.

**Tabla 3** Modelo optimo de red neuronal

$$[P]^* = 1.54h_1 + 0.37h_2 - 0.78h_3$$

$$[GA]^* = -0.69 - 1.74h_1 + 1.89h_2 + 1.22h_3$$

$$h_1 = \frac{1}{1 + \exp(3.69 - 0.07 \hat{a}_m^* - 3.54 \hat{s}^*)}$$

$$h_2 = \frac{1}{1 + \exp(0.74 - 3.71 \hat{a}_m^* + 1.5 \hat{s}^* - 0.38 \hat{t}_m^*)}$$

$$h_3 = \frac{1}{1 + \exp(0.8 - 1.93 \hat{a}_m^* + 7.05 \hat{s}^* + 2.3 \hat{t}_m^*)}$$

2.- (2.5 puntos) Un médico con experiencia está construyendo un sistema de razonamiento basado en casos para automatizar una tarea de diagnóstico. Los casos se corresponden con personas individuales, donde sus datos se componen de una serie de características que describen los posibles síntomas y la parte de solución representa el diagnóstico (clasificación de la enfermedad).

Entrenamiento	Fiebre	Vómitos	Diarrea	Temblores	Clasificación
C1	alta	si	si	no	Into. por Salmonella, (IS)
C2	media	no	si	no	Into. por Salmonella, (IS)
C3	no	si	si	no	Infección intestinal, (II)
C4	media	si	no	no	Infección intestinal, (II)
C5	no	no	si	si	Infección intestinal, (II)
C6	alta	si	si	si	Into. por Salmonella, (IS)
C7	media	no	no	si	Apendicitis (A)
C8	alta	no	no	no	Apendicitis (A)

Construya el árbol de decisión y clasifique un patrón con valores de las características  $q =$  (media; no; no; si).