

IMPLEMENTACIÓN DE REDES NEURONALES PARA LA DETECCIÓN DE LA PRESENCIA DE ENFERMEDADES EN EL CORAZÓN

IMPLEMENTATION OF NEURAL NETWORKS TO DETECT THE PRESENCE OF DISEASE IN THE HEART

Abstract

Then by means of two methods for implementing neural networks and a database that has different characteristics of patients, some have some type of heart disease and others do not. We developed a supervised neural network first part, specifically a multilayer perceptron is implemented, the second part presents unsupervised networks, implementing a network ART2 (Adaptive resonance theory).

Keywords: Epochs, number of neurons, multilayer perceptron, ART networks, learning rate, rate monitoring.

Resumen

A continuación por medio de dos métodos para la implementación de redes neuronales y una base de datos que presenta diferentes características de pacientes, algunos presentan algún tipo de enfermedad del corazón y otros no. Se desarrollo en una primera parte redes neuronales supervisadas, específicamente se implementa un perceptrón multicapa; la segunda parte plantea redes no supervisadas, implementando una red ART2 (Adaptive resonance theory).

Palabras clave: Épocas, número de neuronas, perceptrón multicapa, redes ART, tasa de aprendizaje, tasa de vigilancia.

Jose A. Avellaneda González

Ingeniero Electrónico. Investigador de la Universidad Distrital Francisco José de Caldas. Bogotá, Colombia.
jaavellanedag@correo.udistrital.edu.co

Cynthia M. Ochoa Rey

Ingeniera Electrónica. Investigadora de la Universidad Distrital Francisco José de Caldas. Bogotá, Colombia.
cmochoar@correo.udistrital.edu.co

Tipo: Artículo de revisión

Fecha de Recepción: Sept. 2 de 2010

Fecha de Aceptación: Nov. 5 de 2010

1. INTRODUCCIÓN

En países como Colombia donde se atraviesa por una crisis en los servicios de salud, donde los pacientes no pueden ser atendidos con el tiempo suficiente para curar las enfermedades o diagnosticarlas

a tiempo, es importante tener una herramienta automática que le ayude al médico a diagnosticar y así priorice la atención de los pacientes; por otro lado un incentivo del desarrollo de este trabajo es que en Colombia la principal casusa de muerte son las enfermedades cardiovascu-

lares y lo que se propone en este trabajo puede colaborar con el diagnóstico de anomalías en el corazón que pueden convertirse en problemas más graves de salud

Específicamente se plantea a lo largo de este documento una herramienta basada en redes neuronales supervisadas y no supervisadas, implementadas como perceptrón multicapa o como una red ART2; de estos dos métodos se expondrán las características de cada uno y los resultados obtenidos, para finalmente compararlos.

2. BASE DE DATOS

Las bases de datos originales fueron creadas en julio de 1988 y contiene datos de cuatro sitios diferentes: Fundación Clínica de Cleveland, Instituto Húngaro de Cardiología, Centro Médico de Long Beach y Hospital Universitario de Zurich. Ésta originalmente consta de 76 atributos (incluyendo el atributo a predecir), pero fue reducida a 14 atributos utilizados en otros experimentos con buenos resultados en el diagnóstico, esta base de datos está publicada en la página de la Universidad de Massachusetts Amherst por el centro de aprendizaje automático y sistemas inteligentes.

Respecto a la base modificada, que es la empleada en este trabajo, contiene 270 instancias correspondientes a pacientes de las anteriores clínicas mencionadas, 150 de ellas pertenecen a personas enfermas y 120 a personas sanas. Los atributos que contiene la base de datos pueden ser de tipo real, binario o nominales, de esta forma se clasifican y explican a continuación:

- Atributos con tipo de dato real

Edad: en años, entre 29 años y 77 años.

Presión arterial en reposo: medido en mg/dl (miligramo por decilitro).

Colesterol sérico: medido en mg/dl.

Depresión del segmento ST inducida

por el ejercicio en relación con reposo: dado por el electrocardiograma.

Número de vasos principales: de 0 a 3. Los vasos obstruidos o que presentan anomalías son coloreados por fluoroscopia, por medio de un examen invasivo conocido como angiografía.

- Atributos con tipo de dato nominal

Tipo de dolor en el pecho: valor nominal, de 1 (menor dolor) a 4 (mayor dolor).

Resultados del electrocardiograma en reposo: valor nominal, se califica con 0, 1 y 2.

Pendiente máxima del segmento ST durante el ejercicio: calificado como 1 pendiente ascendente, 2 pendiente 0 y 3 pendiente descendente.

Estado del tejido cardiaco: caracterizado como 3 si es dañado, 6 si tiene daño reversible o 7 si es normal.

Enfermedad: es el atributo que se quiere determinar, 1 representa ausencia de alguna enfermedad o 2 si hay presencia.

- Atributos con tipo de dato binario

Sexo: valor binario, 1 si es hombre y 0 si es mujer.

Glucemia en ayunas: referido a mayor de 120mg/dl.

Angina de pecho inducida por ejercicio: dolor en el pecho. 1 si se presenta, 0 si no se presenta.

3. PREPROCESAMIENTO DE DATOS

3.1 Normalización de los datos

Dependiendo de la red neuronal que se emplea se debe realizar un pre procesamiento a todos los datos para dejarlos entre los valores 0 y 1 o -1 y 1, si no se realiza esta normalización la red neuronal no podrá discriminarlos. Se hicieron cuatro normalizaciones diferentes, como se describen a continuación:

- Normalización 1: Cada dato se divide sobre el máximo del atributo al que co-

responde.

$$X_{iN} = \frac{X_i}{\max(X_{ia})} \quad X_{iN} \in [0, 1]$$

- Normalización 2: Cada dato se divide sobre el máximo del atributo al que corresponde y se le resta la media del conjunto de datos del atributo

$$X_{iN} = \frac{X_i}{\max(X_{ia})} - \text{media}(X_{ia}) \quad X_{iN} \in [-1, 1]$$

- Normalización 3: Cada dato se divide sobre el máximo de todo el conjunto de datos.

$$X_{iN} = \frac{X_i}{\max(X)} \quad X_{iN} \in [0, 1]$$

- Normalización 4: Cada dato se divide sobre el máximo de todo el conjunto de datos y se le resta la media.

$$X_{iN} = \frac{X_i}{\max(X)} - \text{media}(X) \quad X_{iN} \in [-1, 1]$$

De las anteriores ecuaciones X_{iN} es el dato normalizado, X_i es el dato original de la base de datos, X_{ia} es el conjunto de datos por atributo y X es el conjunto de datos completo de la base de datos.

3.2 Análisis PCA

También se realizó un análisis de PCA (Principal Component Analysis) a la base de datos para reducir el número de entradas a las redes neuronales.

Primero se observó la correlación de los atributos de la base de datos de forma gráfica como muestra la Fig. 1, donde los colores cálidos representan el máximo y los colores fríos el mínimo de correlación.

La diagonal en rojo corresponde a la correlación de cada atributo con el mismo; los dos cuadros en naranja es la máxi-

ma correlación con un valor de 0.6 y hace referencia a los atributos de la pendiente máxima y la depresión del segmento ST. De estos resultados podemos decir que no hay atributos prescindibles y que posiblemente PCA no entregará buenos resultados por las propiedades de la base de datos.

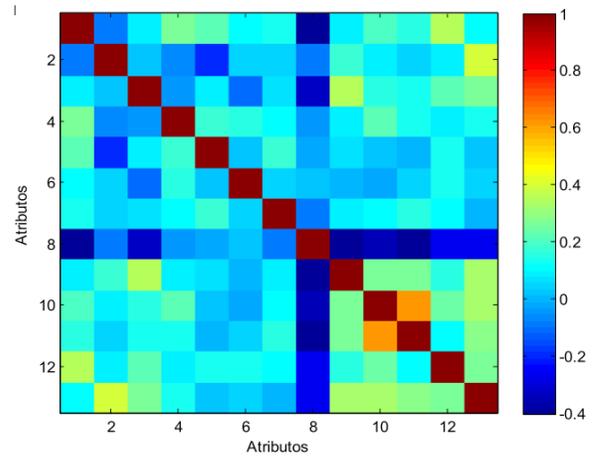


Fig. 1. Correlación entre atributos de la base de datos.

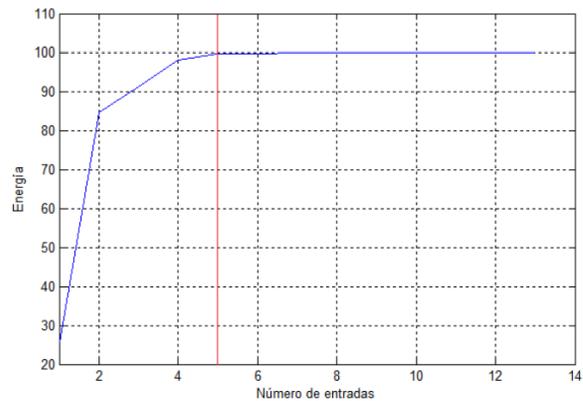


Fig. 2. Curva de energía.

La línea roja muestra que a 5 entradas ya se obtiene aproximadamente un 99% de información, de esta manera se van a realizar pruebas con este número de entradas y con el nuevo espacio de datos. De acuerdo con el error cuadrático medio se tiene un error de 2.58, al realizar esta disminución de atributos.

4. PERCEPTRÓN MULTICAPA

El perceptrón multicapa o MLP es una red neuronal artificial formada por múltiples capas, que le permite resolver problemas que no son linealmente separables; esta es una de las ventajas con respecto a un perceptrón sencillo.

Las redes se crearon con el toolbox de Matlab® nntool, y se realizaron unas primeras pruebas cambiando los parámetros del perceptrón multicapa MLP, por ejemplo, se varían las neuronas entre 8 y 25, para una red con una capa oculta y una de salida; también se cambió la función de activación en diferentes variaciones, la primera capa con tangente hiperbólica, la segunda con logística, al contrario y utilizando ambos tipos de funciones en cada capa. Al colocar tres capas, se emplearon el mismo número de neuronas en la primera capa oculta que en el experimento anterior, en la segunda capa oculta se colocaron neuronas de 3 a 10 neuronas.

De acuerdo con estas primeras pruebas se escogió la siguiente estructura de la red:

- Número de capas: 2
- Número de neuronas: de 13 a 20.
- Número de épocas: 1000

Los parámetros escogidos dependiendo de los aciertos que la red alcanza, para la red con dos capas con neuronas menores a 13, el porcentaje de aciertos es bajo, para más de 20 neuronas no se obtienen mejores porcentajes de aciertos que con menos neuronas; de forma similar los resultados con 3 capas, no son mejores que con sólo dos capas, y si implica mayor costo computacional, por este motivo adicional no se empleó. Por último la tangente hiperbólica en ambas capas, permitía una mejor clasificación con

respecto a otras combinaciones.

Después de obtener la estructura de la red, y establecer las diferentes normalizaciones, se procede a dividir la base de datos en 60% de los datos para entrenar la red neuronal y el 40% restante para probarla, de esta forma quedan en la parte del entrenamiento 88 pacientes sanos y 74 enfermos, en el test quedan 62 y 46 respectivamente.

- Los parámetros de entrenamiento fueron los siguientes:
- Función de entrenamiento: Gradiente Descendiente (TrainGD).
- Función de desempeño: Error Cuadrático Medio (MSE).

De las primeras pruebas también se pudo observar que aunque por defecto Matlab® establece los pesos y bias aleatoriamente, tienen alguna tendencia y las redes siempre inician con desempeños similares; por esto se adicionó una función que inicializara los pesos de acuerdo a valores que entrega la función rand de Matlab®. Esta función se utiliza cuando se tienen tres valores consecutivos del mismo porcentaje de acierto para reiniciar la red, puesto que puede estar en un mínimo local.

5. ANÁLISIS DE RESULTADOS DEL MLP

Se realizan 20 pruebas por cada cambio de parámetros o combinación entre tipo de normalización, número de neuronas y tasa de aprendizaje para tener un total de 3200 pruebas en total. Este número de pruebas es necesario para contrarrestar la inicialización aleatoria de los pesos y bias. En la siguiente tabla se presentan los resultados obtenidos de cada normalización de acuerdo al número de neuronas:

Tabla 1. Mejores resultados obtenidos por cada normalización y número de neuronas.

	No	Tasa de Aprendizaje	Máximo de Aciertos (%)	Promedio de Aciertos (%)	Desviación Estándar
N1	13	0,1	80,56	76,57	2,83
	14	0,5	81,48	69,77	12,95
	15	0,75	79,63	69,44	4,35
	16	0,5	80,56	76,76	2,69
	17	0,75	85,19	76,67	4,89
	18	0,75	85,19	79,40	3,74
	19	0,3	85,19	77,69	5,36
	20	0,75	85,19	80,46	3,80
	N2	13	0,1	84,26	79,03
14		0,5	86,11	78,56	4,44
15		0,9	85,19	80,51	2,63
16		0,1	84,26	81,85	1,67
17		0,9	87,96	81,25	4,35
18		0,5	85,19	80,97	3,52
19		0,3	87,96	83,66	4,39
20		0,1	87,04	82,31	2,97
N3	13	0,5	85,19	80,09	3,50
	14	0,3	83,33	80,28	3,03
	15	0,5	86,11	80,32	2,67
	16	0,5	86,11	80,51	3,75
	17	0,3	84,26	81,06	2,37
	18	0,9	87,96	78,80	5,25
	19	0,9	86,11	81,76	2,96
N4	13	0,1	85,19	79,81	2,65
	14	0,75	84,26	79,58	3,75
	15	0,75	82,41	79,95	2,68
	16	0,75	85,19	81,67	2,55
	17	0,5	86,11	79,44	4,07
	18	0,9	87,04	81,34	3,03
	19	0,3	86,11	81,53	2,70
PCA N1	13	0,1	81,48	77,64	2,54
	14	0,3	79,63	74,26	3,82
	15	0,9	78,70	74,31	2,20
	16	0,9	81,48	70,46	8,63
	17	0,3	78,70	70,88	3,56
	18	0,5	80,56	74,63	4,55
	19	0,3	81,48	68,98	4,42
	20	0,9	79,63	75,83	3,54

PCA N2	13	0,1	80,56	71,85	8,45
	14	0,1	81,48	76,44	2,21
	15	0,3	78,70	76,30	1,43
	16	0,1	83,33	73,84	8,59
	17	0,75	78,70	72,31	4,62
	18	0,75	79,63	72,13	3,52
	19	0,75	80,56	75,51	3,49
20	0,75	80,56	71,71	8,22	

Los mejores resultados de cada normalización están señalados en color rojo. Los mejores resultados se consiguen con una tasa de aprendizaje de 0.75, según la moda de la columna de la tasa de aprendizaje. La normalización por el máximo del dato de cada entrada (normalización 1), tiene los porcentajes de clasificación más bajos, a excepción de este caso no hay cambios significativos en el porcentaje de clasificación al cambiar la normalización.

Algunas desviaciones estándar están en 12 y 8, lo que significa que las pruebas realizadas tienen diferentes resultados y que están alejados unos de otros, esto justifica y prueba la necesidad de realizar varias pruebas para alcanzar mejores resultados.

6. REDES ART2

El modelo ART (Teoría de resonancia adaptativa) soluciona el dilema de la estabilidad y plasticidad del aprendizaje mediante un mecanismo de realimentación entre las neuronas competitivas de la capa de salida. El funcionamiento básico es la asociación de patrones a clases existentes, cuando esto sucede se mueve el centro del clúster adaptándose al nuevo patrón, si no es posible la asociación crea nuevas clases.

Hay diferentes tipos de redes ART, las ART1 utilizan datos binarios, pero como la base de datos tiene valores reales se utilizan las redes ART2 que manejan este tipo de datos.

Los datos también fueron normalizados, pero la red no aceptaba números negativos, por lo que se utilizó inicialmente las normalizaciones 1 y 3.

Para establecer los parámetros de la red, al igual que para las redes MLP se desarrollan pruebas previas; utilizando la normalización 3, los datos quedan muy juntos y son muy pequeños la red no pudo reconocer ningún patrón con esta normalización y finalmente las pruebas se realizaron con la primera normalización.

Las pruebas iniciales realizaban un barrido de los parámetros de vigilancia y tasa de aprendizaje, donde se varía uno solo de estos parámetros dejando el otro fijo, así se observó que la tasa de aprendizaje puede variar entre 0.1 y 1, pero la vigilancia cambia dependiendo del número de atributos que ingresaban a la red; en la figura a continuación, la línea roja representa un límite de clasificación que ocurre cuando sólo se reconoce una clase

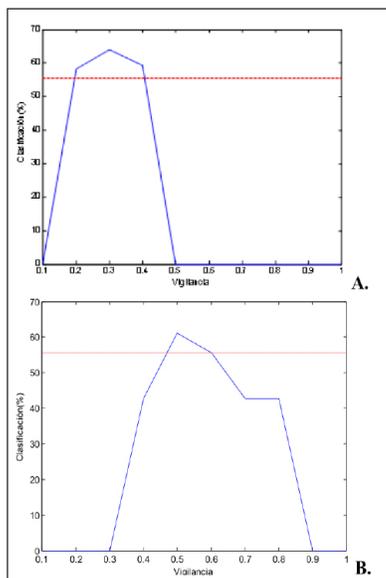


Fig. 3. Variación del porcentaje de clasificación en relación al cambio de parámetro de vigilancia A. Con 13 atributos de entrada a la red neuronal para 0.5 de tasa de aprendizaje. B. Con 5 atributos de entrada a la red neuronal, para 0.1 de tasa de aprendizaje.

Los parámetros finales que se emplearán se resumen en la tabla a continuación:

Tabla 2. Parámetros de entrada para la red ART2.

	13 Atributos de entrada	5 Atributos de Entrada (PCA)
Nº de atributos	26	10
Máx. de categorías	2	
Vigilancia	0.15 0.25 0.35 0.45 0.55	0.45 0.55 0.65 0.75 0.85
Tasa de aprendizaje	0.1 0.3 0.5 0.8 1	

El número de atributos siguen siendo 13 y 5 respectivamente, pero para utilizar las funciones predeterminadas del código, el parámetro de entrada del algoritmo que crea la red ART se coloca 26 y 10; esto es consecuencia de una expansión y reflexión de los datos que es necesario hacer para alejarlos.

Se harán 10 pruebas por cada combinación de parámetros entre vigilancia y tasa de aprendizaje, el número de épocas son 1000 y el criterio de parada del algoritmo son las épocas o cuando no hay actualización en los pesos.

De las pruebas iniciales se observa que la red representa los conjuntos de clasificación con 1, 2 o -1, dependiendo de la vigilancia, por esto se implementó un algoritmo que agrupa estos conjuntos de acuerdo al mejor resultado que se adapte a los datos de entrenamiento, y éste se aplicará a los datos de test para determinar las clases de interés (pacientes sanos o enfermos).

7. ANÁLISIS DE DATOS DE LA RED ART2

Las siguientes tablas consignan los resultados obtenidos en las pruebas con 13 y 5 entradas a la red neuronal.

Tabla 3. Resultados con 13 atributos de entrada a la red neuronal.

Vigilancia	LR	Máxima clasificación [%]	Media	Desviación estandar
0.15	0.10	65.74	15.00	24.99
	0.30	62.04	27.69	29.63
	0.50	70.37	16.94	28.06
	0.80	64.81	23.61	30.65
	1.00	58.33	35.19	25.22
0.25	0.10	57.41	49.54	7.35
	0.30	62.04	54.81	7.09
	0.50	69.44	59.17	7.11
	0.80	62.96	55.19	3.91
	1.00	61.11	57.31	2.49
0.35	0.10	56.48	46.94	16.97
	0.30	63.89	53.33	19.10
	0.50	62.96	54.81	19.36
	0.80	62.04	55.93	7.23
	1.00	72.22	55.83	9.99
0.45	0.10	57.41	57.04	0.89
	0.30	58.33	55.74	4.78
	0.50	59.26	55.83	4.78
	0.80	60.19	53.43	7.60
	1.00	60.19	57.50	1.92
0.55	0.10	60.00	11.83	24.95
	0.30	56.48	15.56	25.33
	0.50	59.26	11.76	24.79
	0.80	53.33	21.57	28.20
	1.00	62.04	59.72	1.09

En rojo se señalan los mejores resultados en porcentaje de clasificación, en tres casos se da con tasa de aprendizaje de 1, un resultado que se espera, pues la red aprenderá más si el parámetro es alto, corriendo el riesgo del sobre entrenamiento, sin embargo con tasa de vigilancia de 0.15 y 0.25 las mejores clasificaciones se dan en 0.5. El mejor resultado de toda la tabla fue 72.22% con vigilancia de 0.35 y 1 de tasa de aprendizaje.

La desviación estándar en la mayoría de los casos es grande, es decir, que las pruebas con los mismos parámetros están alejadas una de la otra, nuevamente ratifica

la necesidad de realizar más pruebas.

Tabla 4. Resultados con 5 atributos de entrada a la red neuronal.

Vigilancia	LR	Máxima clasificación [%]	Media	Desviación estandar
0.15	0.10	65.74	15.00	24.99
	0.30	62.04	27.69	29.63
	0.50	70.37	16.94	28.06
	0.80	64.81	23.61	30.65
	1.00	58.33	35.19	25.22
0.25	0.10	57.41	49.54	7.35
	0.30	62.04	54.81	7.09
	0.50	69.44	59.17	7.11
	0.80	62.96	55.19	3.91
	1.00	61.11	57.31	2.49
0.35	0.10	56.48	46.94	16.97
	0.30	63.89	53.33	19.10
	0.50	62.96	54.81	19.36
	0.80	62.04	55.93	7.23
	1.00	72.22	55.83	9.99
0.45	0.10	57.41	57.04	0.89
	0.30	58.33	55.74	4.78
	0.50	59.26	55.83	4.78
	0.80	60.19	53.43	7.60
	1.00	60.19	57.50	1.92
0.55	0.10	60.00	11.83	24.95
	0.30	56.48	15.56	25.33
	0.50	59.26	11.76	24.79
	0.80	53.33	21.57	28.20
	1.00	62.04	59.72	1.09

Con las 5 entradas, se tienen dos ocasiones que el mejor resultado se logra con 0.1 de tasa de aprendizaje. La desviación estándar no es tan grande comparada con la de la prueba con los 13 atributos, pero se obtienen porcentajes de clasificación más bajos. El porcentaje de clasificación más alto es de 64.81% con vigilancia de 0.45 y 1 de tasa de aprendizaje.

8. COMPARACIÓN ENTRE MLP Y ART2

Para realizar la comparación se compri-

mieron las tablas presentadas en cada método mostrando los mejores resultados por normalización en el caso del MLP o por número de entradas a la red neuronal para ART2.

Tabla 5. Mejores resultados de clasificación por normal.

	N° Neuronas	Tasa de aprendizaje	Máximo de aciertos (%)	Promedio de aciertos (%)	Desviación estándar
N1	18	0,75	85,185	79,398	3,74
N2	19	0,3	87,963	83,657	4,392
N3	18	0,9	87,963	78,796	5,251
N4	18	0,9	87,037	81,343	3,031
PCA N1	13	0,1	81,481	77,639	2,54
PCA N2	16	0,1	83,333	73,843	8,589

Tabla 6. Máximos errores que presenta la red MLP por normalización.

	N° Neuronas	Tasa de aprendizaje	Máximo de aciertos (%)	Promedio de aciertos (%)	Desviación estándar
N1	14	0,5	57,407	30,231	12,950
N2	19	0,1	38,889	24,583	9,316
N3	13	0,1	42,593	25,278	8,317
N4	13	0,5	32,407	22,963	3,861
PCA N1	15	0,1	58,333	36,991	16,969
PCA N2	18	0,5	58,333	35,880	16,019

Tabla 7. Mejores porcentajes de clasificación de la red ART2 por número de entradas.

	Tasa de vigilancia	Tasa de aprendizaje	Máximo de aciertos (%)	Promedio de aciertos (%)	Desviación estándar
13 entradas	0,15	0,50	70,370	16,944	28,058
5 entradas	0,45	1	64,815	58,241	2,877

Tabla 8. Máximos errores de la red ART2 por número de entradas.

	Tasa de vigilancia	Tasa de aprendizaje	Máximo de aciertos (%)	Promedio de aciertos (%)	Desviación estándar
13 entradas	0,15	0,5	100	88,24	24,79
5 entradas	0,45	0,1	100	100	0

De la red MLP, realizando la moda de los mejores resultados por normalización

presentados en la tabla 5, se tiene que ocurren con 18 neuronas y 0.9 de tasa de aprendizaje, con la moda de la tabla 6, se tiene que los mayores errores en clasificación están con 13 neuronas y 0.5 de tasa de aprendizaje. El promedio del porcentaje de aciertos es 83% aproximadamente, 5 puntos porcentuales alejados del máximo, esto significa nuevamente que se deben hacer muchas pruebas para llegar a mejores resultados.

En cuanto a las redes ART2, calculando la moda para las tablas 7 y 8, se tienen que los mejores resultados se encuentran con tasa de aprendizaje de 1 y los mayores errores en clasificación están con tasa de aprendizaje 0.1. Si no se determina el parámetro de vigilancia de forma adecuada se puede tener 100% de error sin importar el número de entradas a la red, como se evidencia en la tabla 8.

Tanto en MLP como en ART2, los porcentajes de clasificación son menores si se utilizan 5 entradas (dadas por PCA).

9. CONCLUSIONES

De los resultados del MLP, se evidencia que los mejores resultados se obtienen con una tasa de aprendizaje de 0.9, casi independiente del número de neuronas y de la normalización empleada.

Para la red ART2, la clasificación de los datos por la red neuronal depende significativamente de la normalización, e inclusive puede no clasificar si se utilizan valores de entradas negativos.

A diferencia de las redes MLP, que no tenían cambios muy grandes al cambiar los parámetros, la red ART2 se ve seriamente afectada por el parámetro de vigilancia, el cuál varía dependiendo de los datos de entrada presentados.

Las redes MLP tienen mejores resultados que las redes ART2, por su porcentaje

máximo de clasificación de 70.3% se puede decir que este método no es el mejor para la solución del problema presentado. Finalmente con PCA no se obtienen me-

jores resultados, como era de esperarse al observar la correlación de las variables; este método de reducción de entradas no es el apropiado para el problema.

Referencias Bibliográficas

- [1] UCI Machine Learning Repository, Statlog (Heart) Data Set [En línea], Disponible: <http://archive.ics.uci.edu/ml/datasets/Statlog+%28Heart%29>
- [2] S. Grossberg, ART (RNA). [En línea], Disponible: <http://cns-web.bu.edu/Profiles/Grossberg/>
- [3] Angiografía por catéter. [En línea], Disponible: <http://www.radiologyinfo.org/sp/info.cfm?pg=angiocath>