SISTEMAS DE INTELIGENCIA ARTIFICIAL

TRABAJO PRÁCTICO Nro. 2

REDES NEURONALES

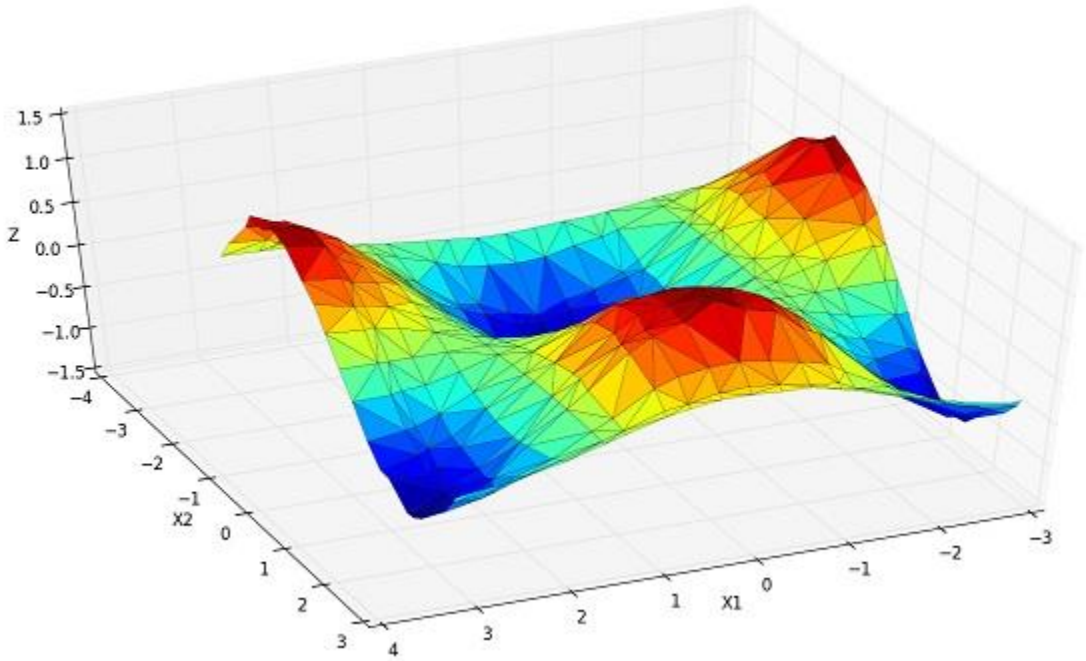
Índice

Objetivo

Crear una red neuronal que aproxime al valor de una función f desconocida. La misma, describe un terreno 3D.

Se deben realizar pruebas con distintas configuraciones de la red y decidir cuál es la que mejor se adapta al problema presentado.

Terreno asignado



Implementación

Las funciones de activación que se implementaron son la exponencial y la tangente hiperbólica. La exponencial tiene un dominio en [0,1], mientras que la tangente tiene su dominio en [-1,1].

Debido a que nuestro terreno posee puntos fuera de dichos rangos, debemos normalizar las entradas y salidas para poder trabajar. Para esto, se aplican las siguientes funciones:

Para la tangente hiperbólica $$\frac{X_i}{|\,Max(x,y,z)|}$$

Para la exponencial $$\frac{sign(X_i) * X_i}{|\,Max(x,y,z)|}$$

Una vez realizada la normalización, se procede con el procesamiento de los patrones.

Backpropagation se implementó de manera batch, por lo que la red utiliza todos los patrones ingresados a ella para el entrenamiento.

Una vez que se procesaron todos los patrones se calcula el error cuadrático medio mediante la ecuación:

$$\frac{\sum_{i\mu}\,(S_I^{\mu} - O_I^{\mu})^2}{2N}$$

Donde N es la cantidad de patrones.

Se considera que la red entrenó lo suficiente una vez que su error cuadrático medio está por debajo de aquel que se considere aceptable en dicha corrida.

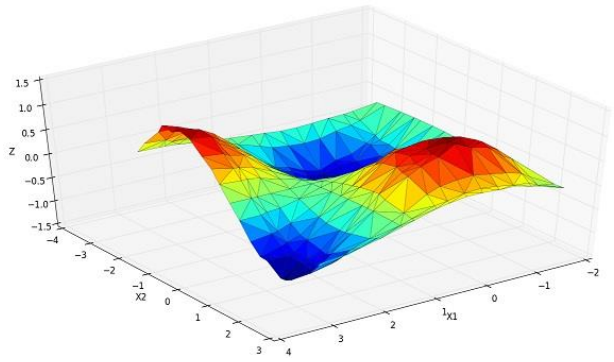
Una vez que se entrenó, se transforman las salidas obtenidas aplicando la inversa de la función de normalización y las matrices de pesos finales se utilizan para realizar la aproximación de la función f desconocida, con los patrones de testeo.

Se finaliza el proceso al obtener el porcentaje de aciertos y de aproximaciones, en base a lo obtenido con los patrones de testeo.

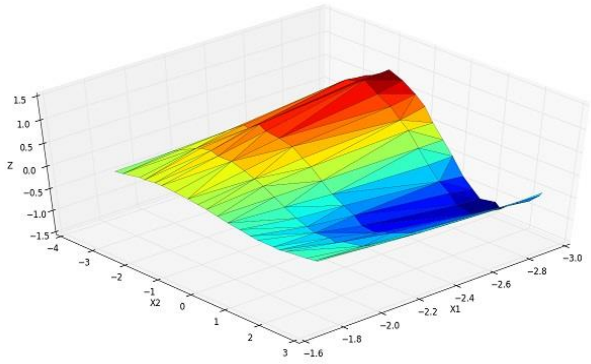
Se considera acierto si

$$S - O < ECM$$

Donde S representa a la salida esperada, O a la salida obtenida y ECM es el error cuadrático medio.



Patrones de entrenamiento



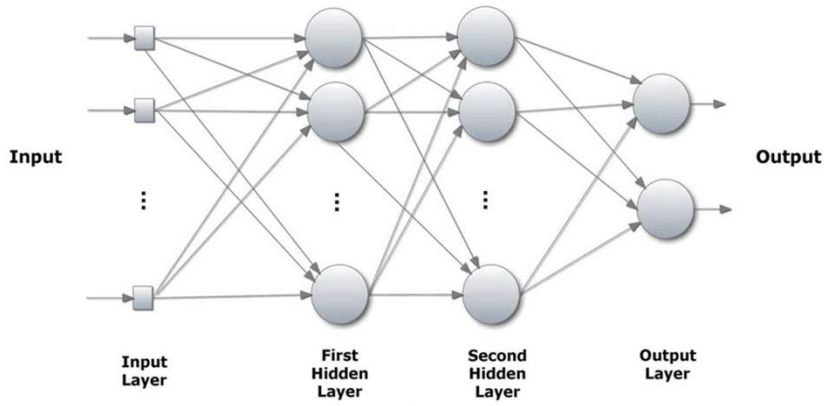
Patrones de testeo

Análisis

Arquitecturas

Para dar comienzo al análisis de nuestra red, realizamos pruebas para determinar cuál era la arquitectura que tenía mejor rendimiento. Consideramos que a mayor porcentaje de aciertos, mayor es el rendimiento de la red.

Consideramos hacer las pruebas con dos capas ocultas desde el comienzo.



En un principio realizamos pruebas con diferencias de cinco neuronas para ambas capas.

Parámetros: ECM = 0.0005 – Eta = 0.5 – Bias = -1 – Beta = 0.5 – Fun = tan

Se utilizó la tangente hiperbólica como función activación debido a que mientras realizamos pruebas para comprobar el correcto funcionamiento de nuestro código, descubrimos que dicha función convergía más rápido el ECM y era más acertada.

En base a los resultados obtenidos (Ver el cuadro de resultados nro. 1 de la sección “Arquitecturas” del anexo), se fueron realizando ajustes para ver si existían variantes que ayudaran a mejorar el porcentaje de aciertos (Ver el cuadro de resultados nro. 2 de la sección “Arquitecturas” del anexo).

Por último, elegimos aquellas que mostraban un mayor rendimiento y repetimos las pruebas cambiando la función de activación por la exponencial (Ver el cuadro de resultados nro. 3 de la sección “Arquitecturas” del anexo).

En base a estos resultados, realizamos los siguientes análisis con dos arquitecturas:

- 2 - 5 -10 - 1
- 2 - 10 - 4 - 1

(# neuronas capa de entrada - # neuronas capa oculta uno - # neuronas capa oculta dos - # neuronas capa de salida)

Coeficiente de aprendizaje

Se realizaron pruebas con tres variaciones de eta: 0.3, 0.5 y 0.7. Estas se hicieron utilizando las arquitecturas mencionadas anteriormente con ambas funciones de activación implementadas (exponencial y tangente hiperbólica).

Parámetros: ECM = 0.0005 – Bias= -1 – Beta = 0.5

Los resultados se encuentran en el cuadro de resultados de la sección “Coeficiente de aprendizaje” en el Anexo.

ECM

Para continuar con los análisis de la red, se procedió a variar los valores del error cuadrático medio aceptable. Estas variaciones son 0.0001, 0.0005 y 0.001. Las corridas se realizaron con las arquitecturas candidatas de los puntos anteriores para ambas funciones de activación.

Parámetros: Eta = 0.5 – Bias = -1 – Beta = 0.5

Los resultados se encuentran en el cuadro de resultados de la sección “Error cuadrático medio” en el Anexo.

Variaciones

Las variaciones de backpropagation que se implementaron fueron Momentum y Eta adaptativo.

Momentum

El término adicional en la actualización de los pesos se denomina momentum, es una memoria o inercia y permite que los cambios en el vector de pesos sean suaves porque incluyen información sobre el cambio anterior

Parámetros: ECM = 0.0001 – Eta = 0.5 – Bias = -1 – Momentum = 0.9

Los resultados se encuentran en el cuadro de resultados de la sección “Momentum” en el Anexo

Eta adaptativo

Un valor de eta constante puede ser bueno en algunos momentos del aprendizaje de la red y en otros no. Si el error no disminuye en una cantidad de iteraciones entonces este debe decrementarse, en caso contrario debe ser incrementado. Se han realizado varias pruebas con esta mejora y el valor de incremento y decremento fue determinado empíricamente en cada una de las pruebas. En particular, esta mejora ha funcionado correctamente junto con la función tangente.

Parámetros: ECM: 0.005 – Eta= 0.05 – Bias= -1 - Momentum= 0.9 – A = 0.0001 – B = 0.01

Los resultados se encuentran en el cuadro de resultados de la sección “Eta adaptativo” en Anexo

Conclusiones

Observamos que en casi todas nuestras pruebas, la función de activación tangente resultó ser la de mayor rendimiento. Creemos que esto puede deberse a que dicha función incluye en su dominio a los valores negativos.

A su vez, pudimos comprobar mediante las diversas pruebas que no necesariamente mayor cantidad de neuronas por capa significa un mayor aprendizaje, sino que depende de la combinación para el problema en cuestión. En nuestro caso, las arquitecturas con las que mejor aprendió nuestra red fue con aquellas que incluían una capa de diez neuronas en su estructura.

La correcta elección de los parámetros puede ser decisiva al momento del rendimiento de la red. Si bien nuestras pruebas muestran una rendimiento cercano al del 50%, algunos análisis fueron descartados ya que al hacer las corridas no se llegaba al error cuadrático medio esperado en un tiempo aceptable (llamamos tiempo aceptable dentro de las 3 hs)

En nuestro caso la variación aplicada del momentum con alpha en 0.9 no afectó de manera significativa nuestros resultados.

Las variaciones implementadas si bien en algún que otro caso aislado mostraron un pico en el rendimiento, el mayor porcentaje de rendimiento mayor a un 60% fue en las pruebas de arquitecturas sin utilizar momentum ni eta adaptativo.

Anexo

Arquitecturas

Cuadro de resultados nro. 1

Resultados				
Capas	**Tiempo (segs)**	**Epocas**	**Aciertos (%)**	**Aproximaciones (%)**
2 5 10 1	279.31	327	80.95	19.05
2 10 5 1	76.58	700	65.71	34.29
2 10 15 1	197.55	208	72.38	27.62
2 15 10 1	405.01	336	66.67	33.33

Cuadro de resultados nro. 2

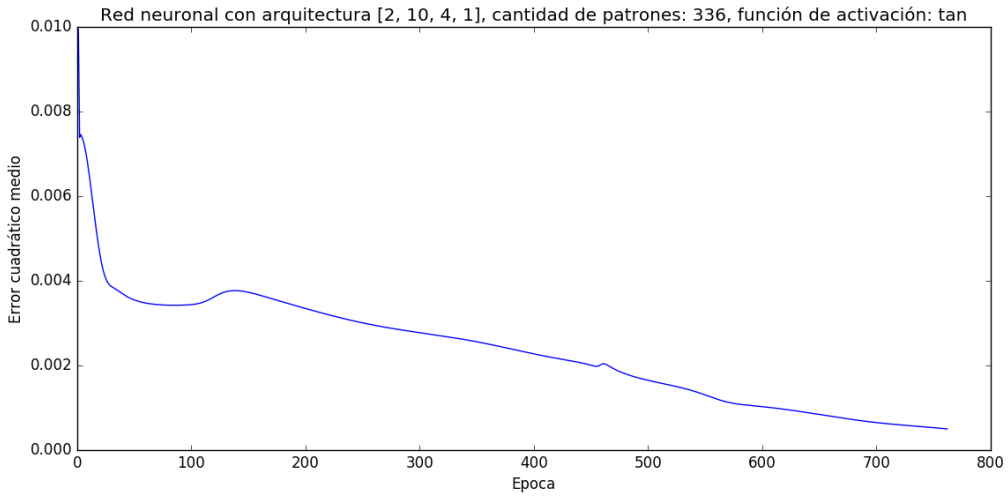
Resultados				
Capas	**Tiempo (segs)**	**Epocas**	**Aciertos (%)**	**Aproximaciones (%)**
2 4 12 1	250.88	341	69.52	30.48
2 8 12 1	584.78	608	73.33	26.67
2 9 3 1	379.38	622	73.33	26.67
2 9 4 1	58.06	554	34.07	62.86
2 8 5 1	360.51	574	71.43	28.57
2 9 6 1	311.3	456	63.81	36.19
2 9 8 1	271.02	359	63.81	36.19
2 12 3 1	440.41	631	65.71	34.29
2 10 4 1	509.25	762	82.86	17.14
2 11 4 1	198.11	289	65.71	34.29
2 12 4 1	76.95	670	57.14	42.86
2 13 4 1	182.52	830	47.62	52.38
2 12 5 1	277.08	367	75.24	24.76

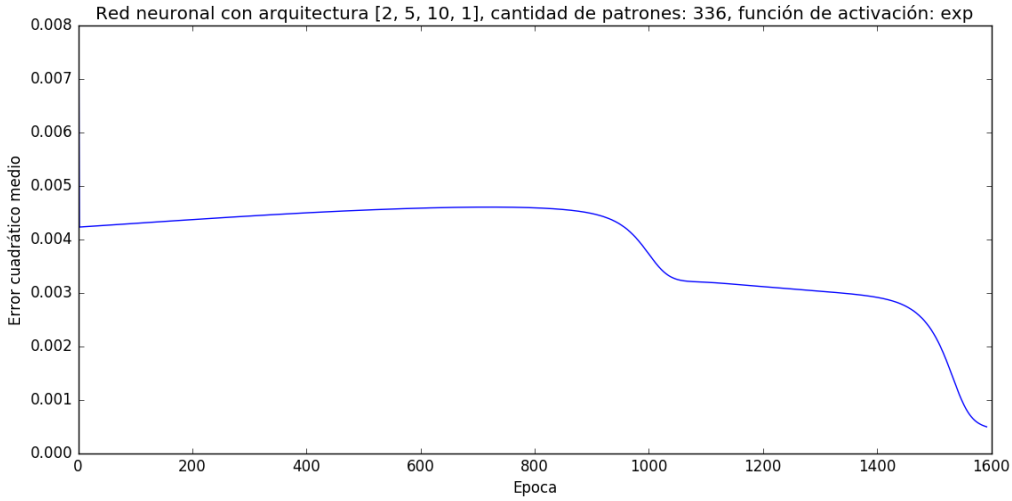
2 10 8 1	286.55	376	64.76	35.24

Cuadro de resultados nro. 3

Resultados				
Capas	**Tiempo (segs)**	**Epocas**	**Aciertos (%)**	**Aproximaciones (%)**
2 5 10 1	1061.33	1592	56.19	43.81
2 10 4 1	1202.34	1871	56.19	43.81
2 12 5 1	516.79	1295	55.24	44.76

ECM – Candidatas





Coeficiente de aprendizaje

Cuadro de resultados

Parámetros			Resultados			
Capas	**Eta**	**Fun**	**Tiempo (segs)**	**Epocas**	**Aciertos (%)**	**Aproximaciones (%)**
2 5 10 1	0.3	exp	221.69	1837	42.68	57.14
2 5 10 1	0.3	tan	111.42	541	32.38	67.42
2 10 4 1	0.3	exp	310.62	2722	44.76	55.24
2 10 4 1	0.3	tan	142.5	717	69.52	30.48
2 10 4 1	0.5	exp	481.55	1394	55.24	44.76
2 10 4 1	0.5	tan	297.7	861	59.05	40.95
2 5 10 1	0.5	tan	134.23	358	69.52	30.48
2 5 10 1	0.5	exp	621.13	1701	55.24	44.76
2 10 4 1	0.7	exp	580.9	908	56.19	43.81
2 10 4 1	0.7	tan	508.12	796	69.52	30.48
2 5 10 1	0.7	exp	1196.02	1809	52.38	47.62
2 5 10 1	0.7	tan	237.35	353	75.24	24.76

Error cuadrático medio (ECM)

Cuadro de resultados

Párametros			Resultados			
Capas	**ECM**	**Fun**	**Tiempo (segs)**	**Epocas**	**Aciertos (%)**	**Aproximaciones (%)**
2 5 10 1	0.0001	exp	474.44	3943	43.81	56.19
2 5 10 1	0.0001	tan	737.66	3553	40.95	59.05
2 10 4 1	0.0001	exp	453.15	3906	43.81	56.19
2 10 4 1	0.0001	tan	6504.26	9984	52.38	47.62
2 10 4 1	0.0005	exp	7704.26	1506	54.19	43,81
2 10 4 1	0.0005	tan	182.44	531	57.14	42.86
2 5 10 1	0.0005	tan	96.35	265	80.95	19.05
2 5 10 1	0.0005	exp	462.99	1276	55.24	44.76
2 10 4 1	0.001	exp	489.33	761	55.24	44.76
2 10 4 1	0.001	tan	247.37	387	50.48	49.52
2 5 10 1	0.001	exp	618.11	915	56.19	43.81
2 5 10 1	0.001	tan	162.47	239	65.71	34.29

Momentum

Cuadro de resultados

Parámetros	Resultados			
Capas	**Tiempo (segs)**	**Epocas**	**Aciertos (%)**	**Aproximaciones (%)**
2 5 10 1	913.89	2615	52.38	47.62
2 5 10 1	75.28	664	67.62	32.38
2 10 4 1	106.43	946	54.29	45.71
2 10 4 1	74.23	373	21.9	78.1

Eta adaptativo

Cuadro de resultados

Parámetros			Resultados			
Capas	**a**	**b**	**Tiempo (segs)**	**Epocas**	**Aciertos (%)**	**Aproximaciones (%)**
2 5 10 1	0.0001	0.01	559.21	1007	60	40
2 10 4 1	0.001	0.001	807.84	7363	56.19	43.81