



Sistemas de Inteligencia Artificial

Trabajo Práctico Especial 2:

Redes Neuronales

Grupo 7:

- | | | |
|--------------------------------|----------------------|-------|
| • María de los Angeles Arlanti | marlanti@itba.edu.ar | 53373 |
| • Mauricio Minestrelli | mminestr@itba.edu.ar | 52015 |
| • Santiago Ocamica | socamica@itba.edu.ar | 53346 |
| • Agop Matías Hurmuz | ahurmuz@itba.edu.ar | 53248 |

Profesores:

- Parpaglione, María Cristina
- Pierri, Alan

Problema

Una empresa de videojuegos precisa un desarrollo que pueda simular en su plataforma terrenos de diferentes partes del mundo a partir de mediciones de altura, latitud y longitud. La red neuronal multicapa que se implemente deberá poder aproximar a la función de altura. Cada grupo debe construir los patrones de entrenamiento a partir del muestreo obtenido.

Objetivo

El objetivo del trabajo es crear una red neuronal multicapa a la cual se la pueda entrenar con un conjunto de patrones para luego generalizar a partir de estos, pudiendo simular terrenos de diferentes partes del mundo.

Introducción

En el informe se explicarán las decisiones tomadas a la hora del armado de la red neuronal. El trabajo fue implementado en Octave.

Al ser una red multicapa genérica admite cualquier tipo de arquitectura, se utilizó eso para poder encontrar la más adecuada para nuestro problema. Los patrones utilizados para el conjunto de entrenamiento como para el de pruebas fueron los otorgados en el archivo terrain07.data provisto por la cátedra.

Para la implementación se respetaron los modelos de de feed-forward y back-propagation, esta última con distintos ajustes. Vamos a considerar una configuración correcta de la red si: el error es menor al error cuadrático medio. También se detallará cómo se eligió la arquitectura y que función de activación se utilizó entre otras cosas.

Toma de decisiones

El conjunto utilizado para ver el terreno a aprender fue otorgado por la cátedra. El mismo fue separado en dos conjuntos disjuntos (cada uno con puntos distintos) uno de entrenamiento y lo restante para testeo. Desde la configuración se puede elegir qué porcentaje se quiere utilizar para el conjunto de entrenamiento, con lo cual lo restante se le otorga a testeo. La elección de puntos es al azar en cada corrida.

Empíricamente se pudo concluir que la combinación 80% de los puntos para entrenamiento y el 20% para testeo fue la que mejor se adaptó al problema. El conjunto de entrenamiento debe ser mayor dado que la red necesita más cantidad de puntos representativos para

aprender, pero no debe ser tan grande (ejemplo 90%) para tener una mejor idea de la generalización de la red.

En cuanto a la función de activación se decidió utilizar la exponencial ya que nuestros puntos de salida (los z) se comprenden en el rango de 0 a 0.9 el cual está completamente incluido en la imagen de dicha función que va de 0 a 1.

Error de Testeo vs Error de Entrenamiento

Pruebas y análisis de resultados

Arquitecturas

Para encontrar la arquitectura con mejor performance para resolver el problema se ideó e implementó un experimento de resolución de problema. Para 13 arquitecturas distintas se intentó aproximar el terreno dado por la cátedra, dejando fija la semilla de los pesos aleatorios para evitar calcular el promedio para una misma arquitectura de red, un mismo β , α y distribución de patrones (80%-20% entrenamiento-testeo) con una cota de error de aprendizaje de 0.001.

Se analizaron arquitecturas de 1, 2 y 3 capas ocultas. Para cada una se varió la cantidad de neuronas en las capas ocultas.

El criterio de corte fue una cota de error de entrenamiento, esta decisión se tomó ya que el terreno al ser para un videojuego requiere ser lo más preciso posible para que su jugabilidad sea mejor, sin importar el tiempo que tarde ya que una vez entrenada la red, se le puede pasar los patrones que se requiere y esta va a responder con la mayor precisión posible de acuerdo al error elegido para entrenar.

Para todas estas arquitecturas se analizó la época en la que aprendió cada una y el error cuadrático medio de testeo, dándole mayor énfasis en el análisis a éste último valor, ya que da una idea de cómo generaliza la red una vez que ésta aprendió.

Algunas de las arquitecturas evaluadas fueron:

- 2, 50, 1
- 2, 10, 1
- 2, 5, 1
- 2,35,10,1
- 2,10,35,1
- 2, 4, 4, 1
- 2, 10, 10, 10, 1
- 2, 30, 10, 30, 1

En el [Anexo 1](#) se encontrará una referencia tabulada de las pruebas para cada arquitectura.

Mejor Arquitectura

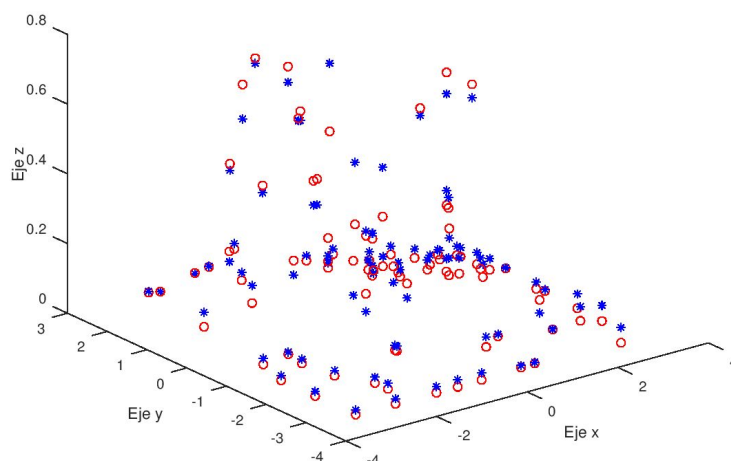
El criterio para elegir la arquitectura que mejor aprendió se basó en elegir aquella que generalizó mejor. Es decir, tomamos en cuenta el menor error cuadrático medio de testeo, sin importar cuánto haya tardado en aprender.

Nos quedamos con la mejor arquitectura para cada grupo (1, 2 o 3 capas ocultas) con mejoras y sin mejoras y luego realizamos una comparación general entre los tres grupos obteniendo así la de menor error cuadrático medio de testeo ([ver comparaciones en Anexo 2](#)).

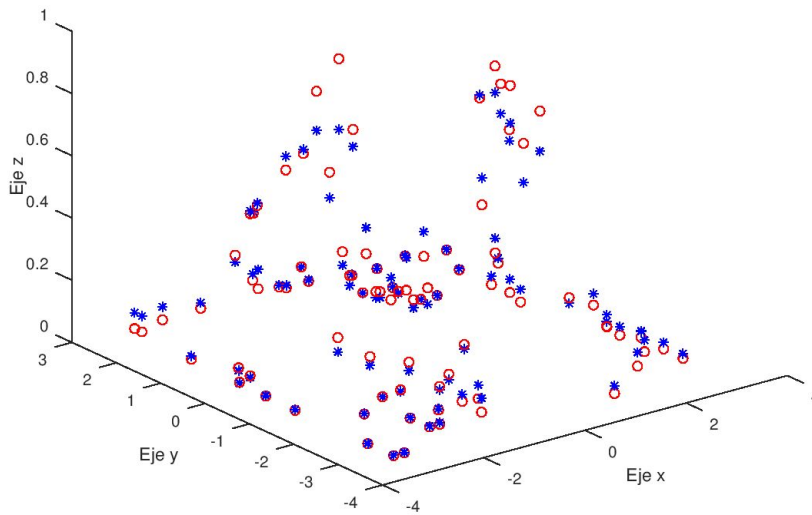
De esta forma, se obtuvo que la mejor arquitectura fue la de 1 capa oculta con la mayor cantidad de neuronas dentro de su grupo, con un orden de 10^{-4} :

Capas	1 Oculta
Arquitectura	2,50,1
Épocas	3986
Tiempo (min)	30.331
Orden Error Testeo	10^{-4}

Se puede observar en los siguientes gráficos que la red generaliza mejor para la arquitectura 2,50,1 que para la 2,10,1.



Generalización arquitectura 2,50,1 con optimizaciones



Arquitectura 2,10,1 con optimizaciones

Mejoras

Una vez determinada empíricamente la arquitectura que mejor se adecuaba al problema entre aquellas probadas se procedió a ajustar el resto de los parámetros.

Variación de β

Luego de encontrar un buen valor de η se probaron distintos valores de β con el fin de encontrar aquel que mejor se adapte a la arquitectura y el problema.

A medida que se incrementa β , la función de activación tiende a la función escalón por lo que su salida da valores menos distinguibles y más cercanos a $|1|$. Por ello, a la red le cuesta más aprender, por el contrario si β fuera pequeño la red no podría distinguir los patrones de forma adecuada.

En base a las pruebas realizadas se optó por utilizar $\beta = 0,5$.

Ajuste de δ

Una de las mejoras implementadas fue el ajuste de la diferencia entre el valor esperado y el obtenido por la derivada de h para garantizar que este producto δ no sea cero. Para lograr esto se sumó el valor 0.1 a la derivada de h :

$$\delta i_{\mu} = [g'(h\mu_i) + 0.1](Si_{\mu} - o\mu_i)$$

En las tablas 1,2,3 vs 4,5,6 se puede ver la comparativa para las arquitecturas probadas entre aquellas en la que se incluye la mejora y aquellas en las que no.

Momentum

La otra de las mejoras implementadas fue la utilización de *momentum* a la hora de la actualización de los pesos. El momentum es un término que pesa el descenso promedio, favoreciendo una oscilación que permita a la red salir de un plateau.

En las pruebas para encontrar el valor adecuado de esta mejora se busca un grado de oscilación suficiente como para sacar a la red de mínimos locales pero sin entrar en una oscilación constante e improductiva que resulte peor que la implementación de la mejora.

Para esto se realizaron pruebas $\alpha=0,9$.

Anexo

Anexo 1: Pruebas de Arquitecturas

Comparación de Arquitecturas 80% patrones de aprendizaje
con $\beta = 0.5$ $\eta = 0.05$ $\alpha = 0.1$ en los casos que corresponda cada parametros en las tablas

Sin mejoras

Tabla: 1 Capa Oculta

Arquitectura	2,50,1	2,10,1	2,5,1
Épocas	3986	4225	4122
Tiempo (min)	30.331	29.198	31.538
Error Cuadrático Medio Testeo	5.6847e-04	7.8323e-04	9.4011e-04

Tabla: 2 Capas Ocultas

Arquitectura	2,35,10,1	2,10,35,1	2,20,5,1	2,4,4,1	2,8,3,1
Épocas	3098	1796	2698	3004	1485
Tiempo (min)	18,987	10,99	15,869	17,792	8,7076
Error Cuadrático Medio Testeo	5,9098E-04	0,0013626	6,6039e-04	0,0012533	6,8452e-04

Tabla: 3 Capas Ocultas

Arquitectura	2,10,10,10,1	2,30,30,30,1	2,10,30,30,1	2,30,10,30,1	2,30,30,10,1
Épocas	4070	5216	2588	1811	5733
Tiempo (min)	44,119	56,388	44,12	14,251	62,367
Error Cuadrático Medio Testeo	0,0017586	0,0015534	0,0014292	0,00093602	0,0011327

Ajuste de Delta (sumado de 0,1)

Tabla 3: 1 Capa Oculta

Arquitectura	2,50,1	2,10,1	2,5,1
Épocas	2609	2972	2361
Tiempo (min)	15,374	17,86	14,461
Error Cuadrático Medio Testeo	0,0011839	8,8682E-04	0,001023

Tabla 4: 2 Capas Ocultas

Arquitectura	2,35,10,1	2,10,35,1	2,20,5,1	2,4,4,1	2,8,3,1
Épocas	2676	1201	2121	1894	3697
Tiempo (min)	36,61	14,08	25,37	24,87	42,42
Error Cuadrático Medio Testeo	0,00127	0,00130	0,00086	0,00131	0,00114

Tabla 5: 3 Capas Ocultas

Arquitectura	2,10,10,10,1	2,30,30,30,1	2,10,30,30,1	2,30,10,30,1	2,30,30,10,1
Épocas	7192	5837	2087	3450	5503
Tiempo (min)	113,34	92,16	33,22	54,64	88,03
Error Cuadrático Medio Testeo	0,00161	0,00176	0,00095	0,00131	0,00069

Momentum

Tabla 6: 1 Capa Oculta

Arquitectura	2,50,1	2,10,1	2,5,1
Épocas	1739	2440	5335
Tiempo (min)	10,420	14,354	31.487
Error Cuadrático Medio Testeo	0,0014273	0,0010082	9,7949E-04

Tabla 7: 2 Capas Ocultas

Arquitectura	2,35,10,1	2,10,35,1	2,20,5,1	2,4,4,1	2,8,3,1
Épocas	2439	1499	3693	1707	2539
Tiempo (min)	34,25	19,49	47,10	22,48	33,96
Error Cuadrático Medio Testeo	0,00055	0,00189	0,00192	0,00104	0,00213

Tabla 8: 3 Capas Ocultas

Arquitectura	2,10,10,10,1	2,30,30,30,1	2,10,30,30,1	2,30,10,30,1	2,30,30,10,1
Épocas	3753	2781	3751	2691	6949
Tiempo (min)	63,95	45,54	65,52	48,30	111,05
Error Cuadrático Medio Testeo	0,00163	0,00118	0,00187	0,00122	0,00065

Delta y Momentum

Tabla 9: 1 Capa Oculta

Arquitectura	2,50,1	2,10,1	2,5,1
Épocas	3466	2135	3164
Tiempo (min)	20,684	12,803	29,141
Error Cuadrático Medio Testeo	4,456E-04	0,001267	9,09E-04

Tabla 10: 2 Capas Ocultas

Arquitectura	2,35,10,1	2,10,35,1	2,20,5,1	2,4,4,1	2,8,3,1
Épocas	2303	1217	3101	2197	2931

Tiempo (min)	18,059	9,2265	23,706	16,697	22,184
Error Cuadrático Medio Testeo	0,001221	0,0011689	0,0010327	0,0021341	0,0013911

Tabla 11: 3 Capas Ocultas

Arquitectura	2,10,10,10,1	2,30,30,30,1	2,10,30,30,1	2,30,10,30,1	2,30,30,10,1
Épocas	4625	3870	1891	2382	6414
Tiempo (min)	44,946	38,106	18,506	23,189	63,379
Error Cuadrático Medio Testeo	0,0021762	0,0010954	0,0011524	0,0010997	0,0016727

Anexo 2: Análisis mejor arquitectura

Tabla 13: Sin mejoras

Capas	1 Oculta	2 Ocultas	3 Ocultas
Arquitectura	2,50,1	2,35,10,1	2,30,10,30,1
Épocas	3986	3098	1811
Tiempo (min)	30.331	18,987	14,251
Error Cuadrático Medio Testeo	5.6847e-04	5,9098E-04	0,00093602

Tabla 14: Con Delta y Momentum

Capas	1 Oculta	2 Ocultas	3 Ocultas
Arquitectura	2,50,1	2,35,10,1	2,30,10,30,1
Épocas	3466	2303	2382
Tiempo (min)	20,684	18,059	23,189
Error Cuadrático Medio Testeo	4,456E-04	0,001221	0,0010997

