Redes Neuronales Multicapa

Sistemas de Inteligencia Artificial

De Santi Pereyra Pintos

24 de Abril de 2013

Tabla de contenidos

Problema propuesto

El problema propuesto fue la estimación de puntos de una serie temporal a partir de una cierta cantidad de puntos anteriores.

$$x_t = f(x_{t-1}, x_{t-2}, \ldots)$$

No se sabe la cantidad de puntos anteriores a x_t que se deben tomar para poder estimar.

Gráfico de la serie temporal

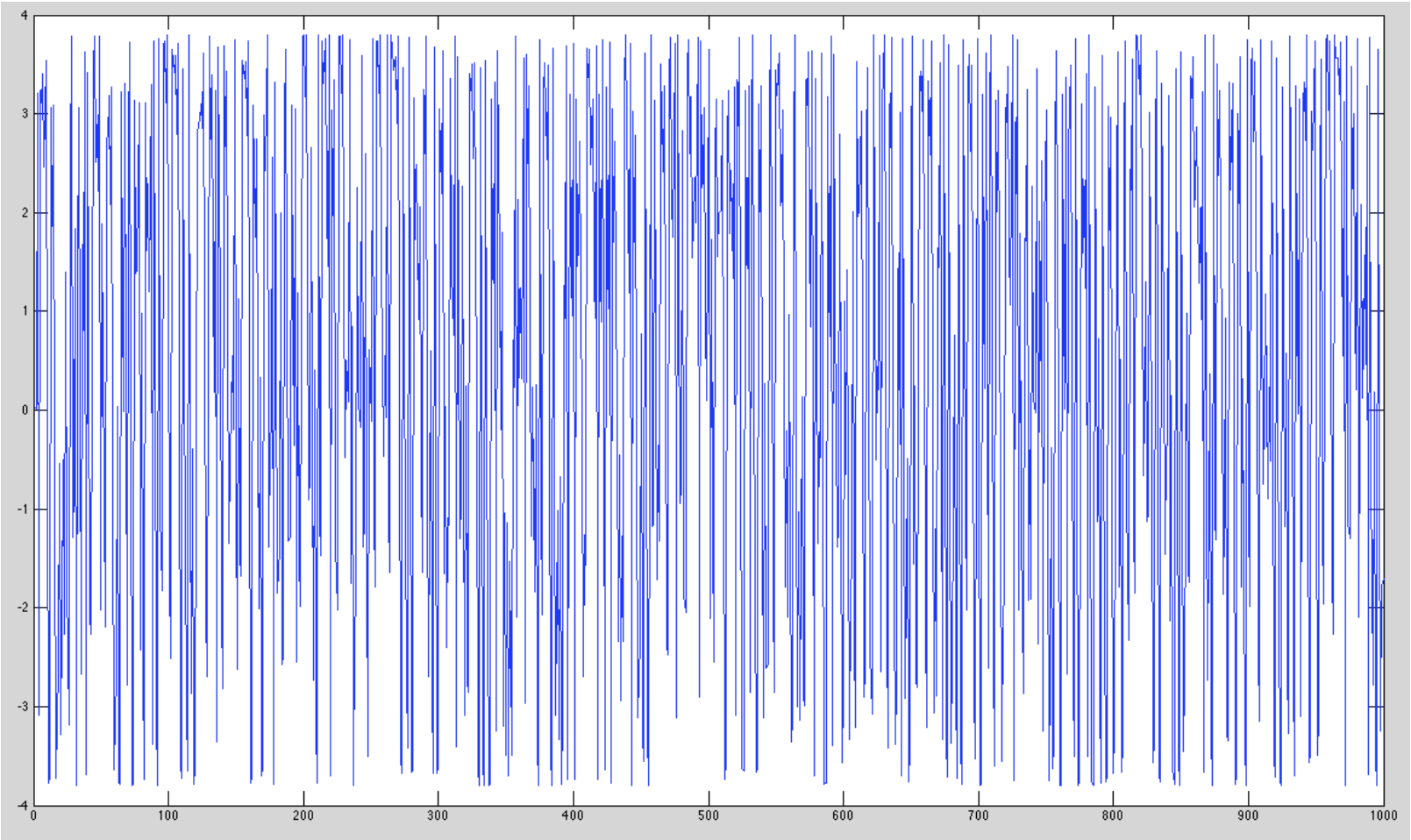


Tabla de contenidos

Representación de la red neuronal

Se representó la red neuronal como una matriz de pesos.

- Cada neurona es una fila de pesos.
- La matriz tiene tantas filas como neuronas haya en la red.

Sigmoidea exponencial

$$f(x) = \frac{2}{1 + e^{-2\beta x}} - 1$$

$$f'(x) = 2\beta \frac{e^{\beta x}}{(e^{\beta x} + 1)^2}$$

Tangente hiperbólica

$$f(x) = tanh()$$

$$f'(x) = \beta sech(x)^2$$

¿Qué arquitectura utilizar?

Se probaron diferentes arquitecturas y en base a los resultados se decidió cuál es la mejor arquitectura.
Se tuvo en cuenta:

- Cantidad de capas mínima
- Cantidad de neuronas mínima
- ¿Error aceptable en el conjunto de entrenamiento?
- ¿Error aceptable en el conjunto de testeo?

Conjunto de entrenamiento y testeo

A partir de los puntos de la serie, se tomo un conjunto al azar para entrenar la red y el resto de los puntos se utilizaron para testear la red entrenada.

Tabla de contenidos

η adaptativo

- Si el error disminuye de manera consistente incrementar η:
  - ¿Cuánto? 0.1
  - ¿Cuántas veces seguidas se considera consistente? 4 veces
- Si el error incrementa, reducir η:
  - ¿Cuánto? A la mitad

Estos parámetros afectan al rendimiento de la red!

Momentum

$$w_{ij}(t+1) = -\frac{\partial E}{\partial wij} + \alpha w(t)$$

Cambios dependen de cambios anteriores. Idea de dirección general del error. ¿Ayuda un α alto?
Se pasan todos los patrones, corrigiendo los pesos en cada caso.
Una vez pasados, se corrige el eta y se vuelve a comenzar.

Ruido en la derivada de la función de activación

Idea: Escapar de mínimos locales.
Se le suma un pequeño delta a la derivada para evitar que valga 0 en los mínimos y así poder evitar mínimos locales.

Tabla de contenidos

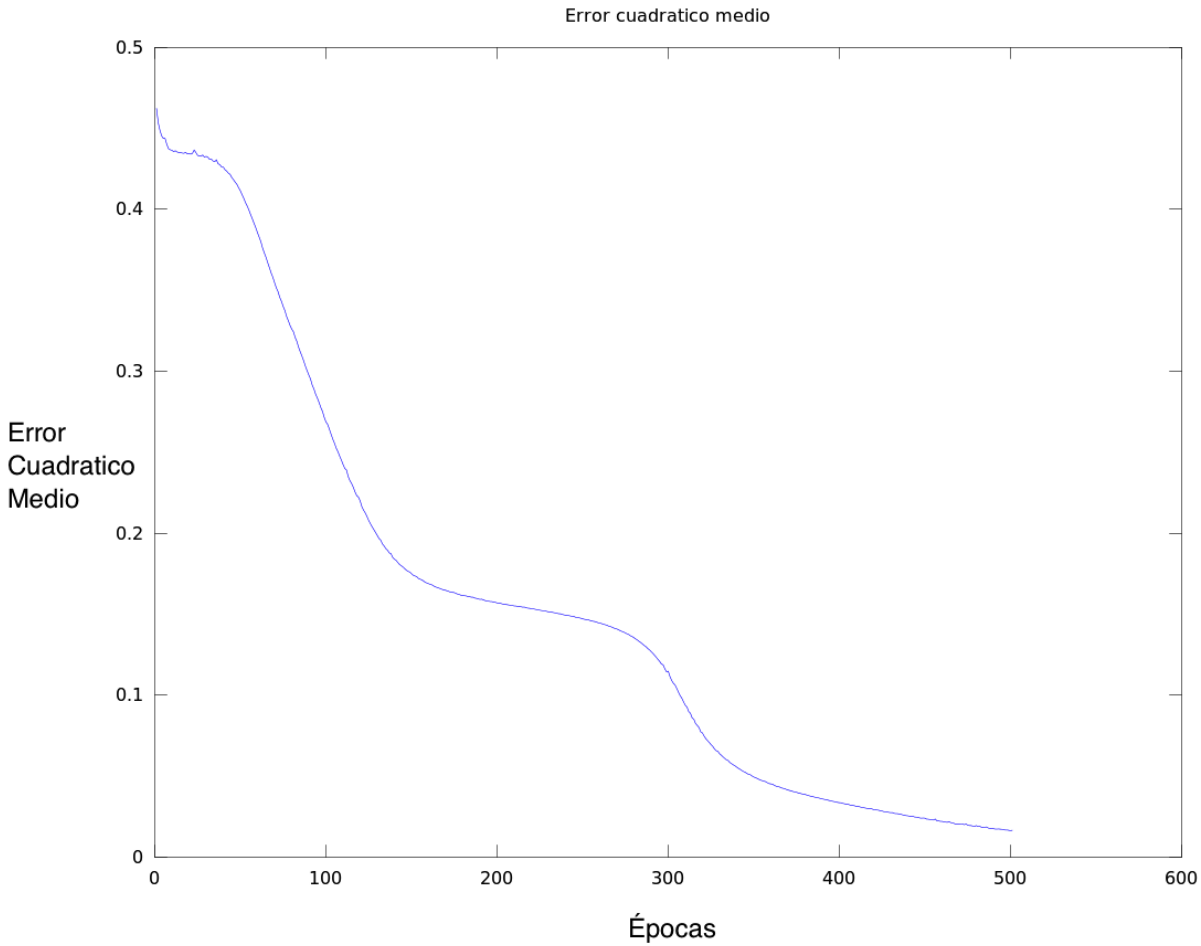
Resultados

- Realizamos pruebas exhaustivas en de distintas arquitecturas de red, conjuntos de testeo, combinaciones de momentum y η adaptativo.
- A continuación mostramos las pruebas más significativas a nuestro criterio.

Primera Prueba

- Para esta prueba utilizamos un η no adaptativo y 500 épocas.
- El tamaño del conjunto de entrenamiento fue de 400 elementos.
- Utilizamos una arquitectura de 20 neuronas en una única capa oculta.
- Obtuvimos un error cuadrático medio de 0,01662 en el conjunto de entrenamiento.
- El error cuadrático medio en el conjunto de prueba fue de 0,22923.
- Con respecto al conjunto de prueba el 66,33 % tuvo un error menor a 0.4.

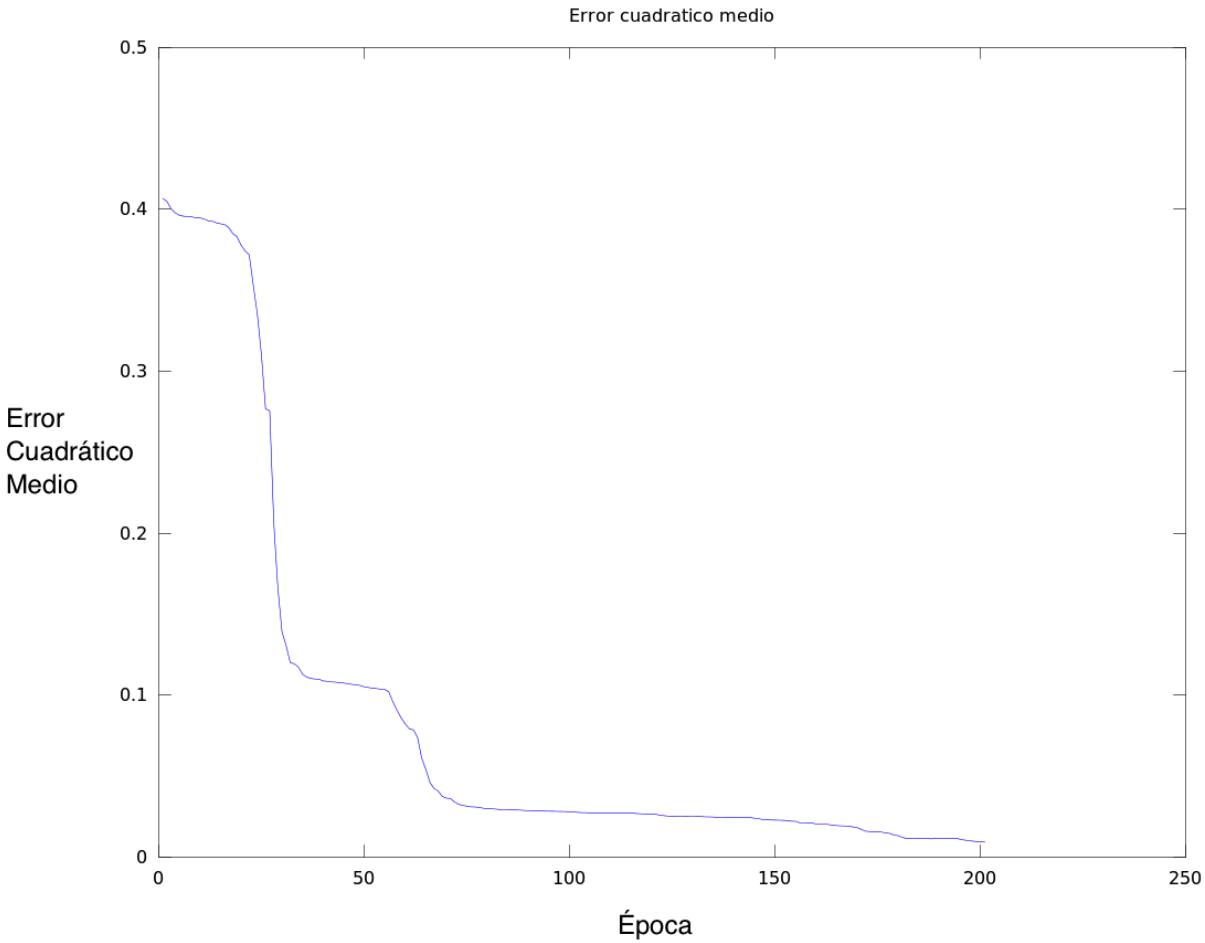
Primera Prueba



Segunda Prueba

- Para esta prueba deshabilitamos el momentum probamos por lo cual corrimos 200 épocas.
- El tamaño del conjunto de entrenamiento fue de 400 elementos.
- Utilizamos una arquitectura de 9 neuronas en la primera capa oculta, y 6 en la segunda.
- Obtuvimos un error de 0,00961 en el conjunto de entrenamiento.
- El error cuadrático medio en el conjunto de prueba fue de 0,17816.
- Con respecto al conjunto de prueba el 74 % tuvo un error menor a 0.4.

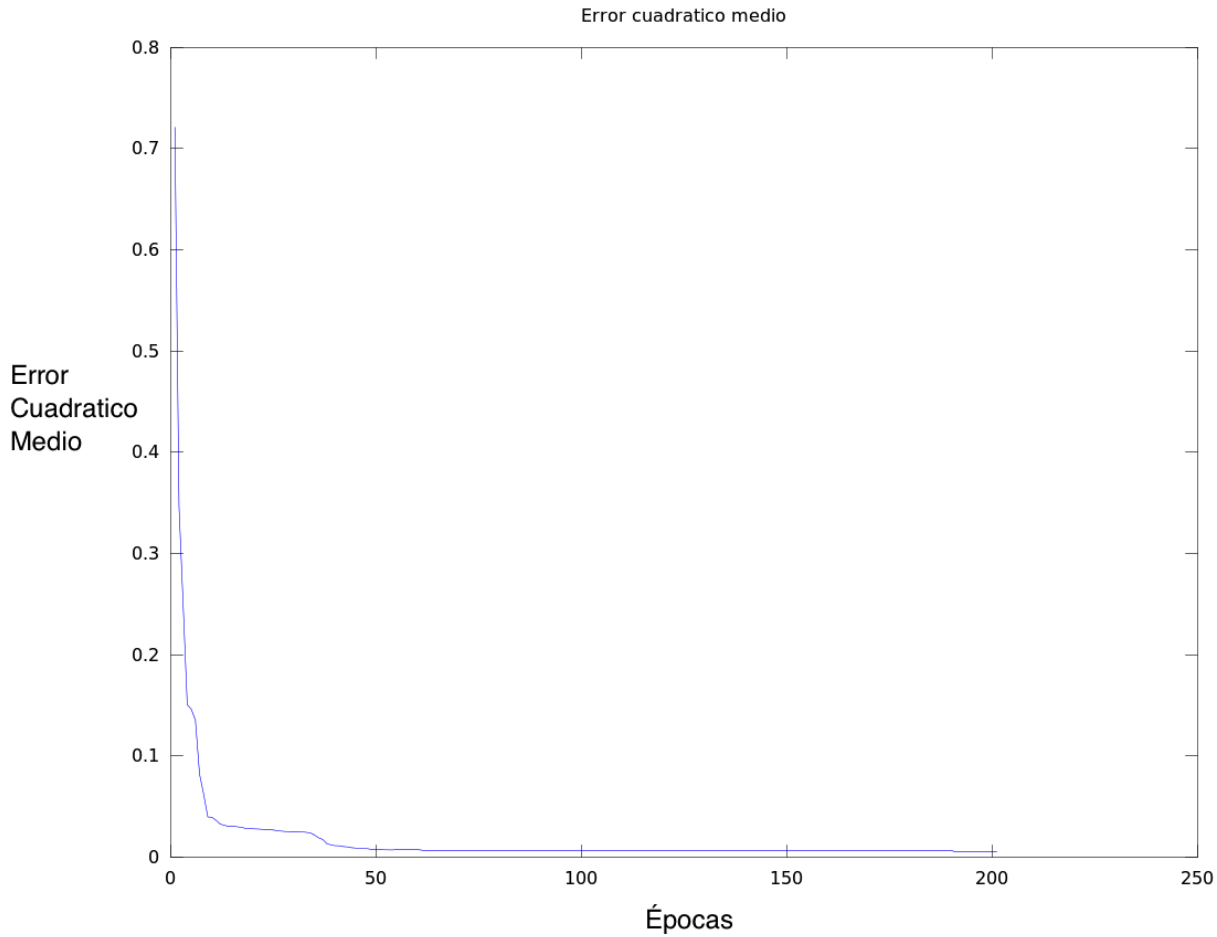
Segunda Prueba



Tercera Prueba

- Para esta prueba habilitamos tanto momentum como η adaptativo.
- El tamaño del conjunto de entrenamiento fue de 800 elementos.
- Utilizamos una arquitectura de 9 neuronas en la primera capa oculta, y 6 en la segunda.
- Obtuvimos un error cuadrático medio de 0,00598 en el conjunto de entrenamiento.
- El error cuadrático medio en el conjunto de prueba fue de 0,07760.
- Con respecto al conjunto de prueba el 83 % tuvo un error menor a 0.4.
- Fue el mejor resultado en nuestra tanda de pruebas.

Tercera Prueba



Tercera Prueba

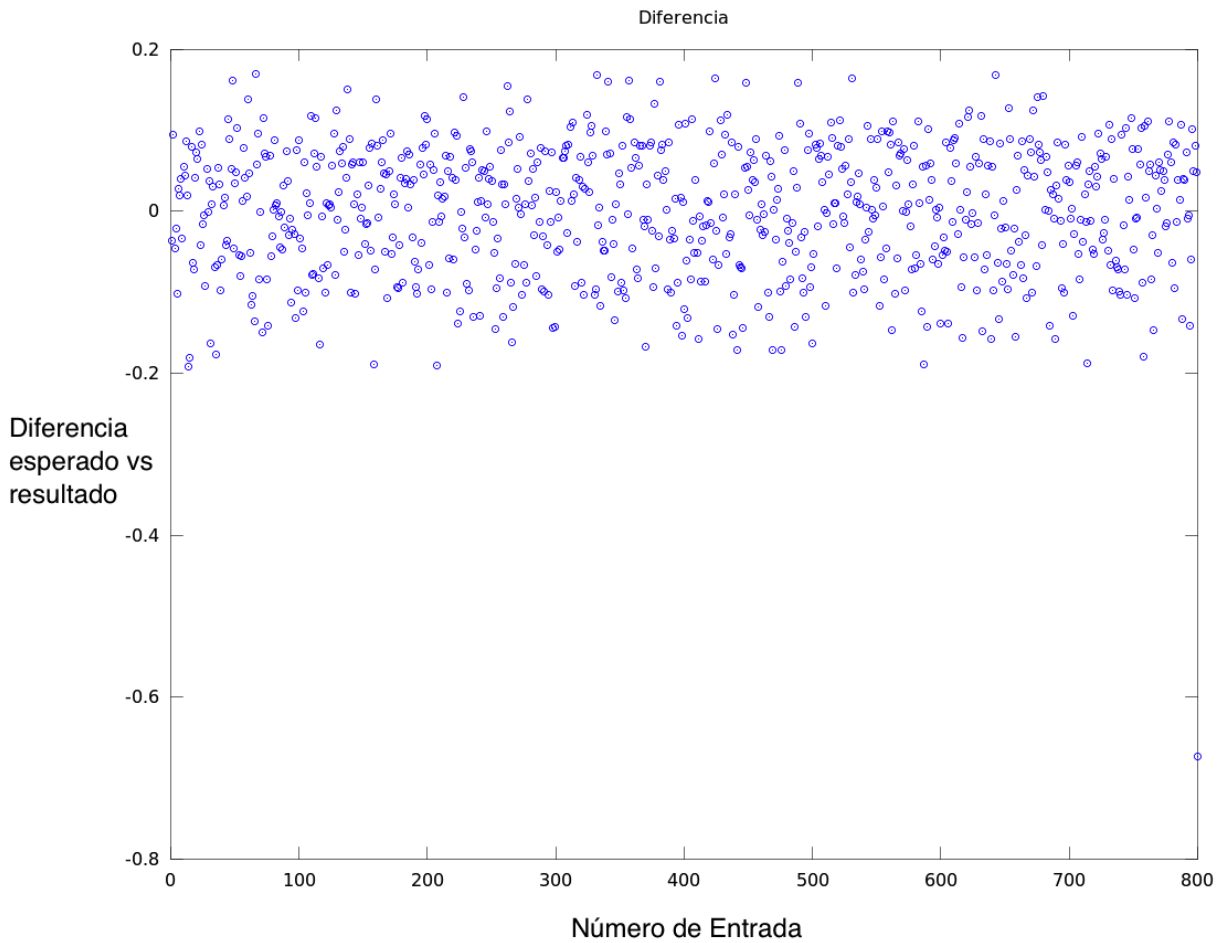


Tabla de contenidos

Conclusiones

- Incrementar la cantidad de neuronas no necesariamente implica mejoras en cuanto al error.
- Los parametros dependen del problema que se esta resolviendo y su correcta elección es de gran importancia.
- Activar el momentum no siempre acelera la convergencia.
- Es posible aprender más aumentando el tamaño del conjunto de entrenamiento.