Redes neuronales multicapa

Castiglione, Karpovsky, Sturla

Sistemas de Inteligencia Artificial

3 de Mayo de 2012

Tabla de contenidos

El problema

El problema planteado consiste en la estimación de funciones escalares a partir de un conjunto de puntos que las representan.

En nuestor caso particular hemos trabajado con el archivo **samples7.txt**

Gráfico de la función

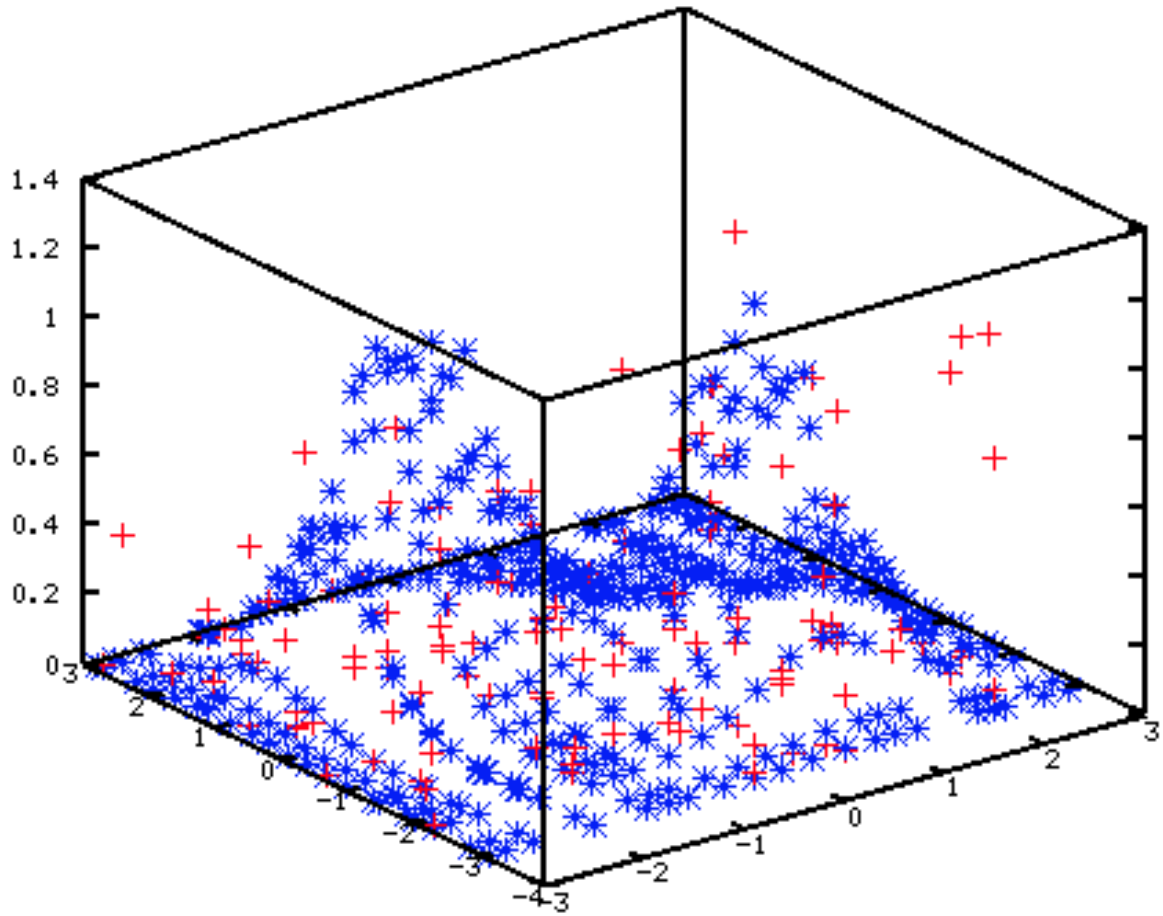

Tabla de contenidos

Representación de la red neuronal

Se representó la red neuronal como una matriz de pesos.

- Cada neurona es una columna de pesos.
- Cada capa de neuronas es una matriz de pesos.
- La red neuronal, por consiguiente, es un vector de matrices.

Se utilizaron dos funciones de activación distintas:

Sigmoidea exponencial

$$g(h) = \frac{1}{1 + e^{-2\beta h}}$$

Derivada:

$$2\beta g(1 - g)$$

Tangente hiperbólica

$$g(x) = tanh(x)$$

Derivada:

$$\beta g(1 - g^2)$$

Arquitecturas estudiadas

- Perceptron simple?
- Pocas neuronas, pocas capas
- Muchas neuronas, muchas capas
- Punto intermedio
- Conexiones muertas/rotas

Conjunto de entrenamiento y testeo

Se decidió seguir el consejo de la cátedra y al realizar las pruebas se utilizó un subconjunto de los datos seleccionados al azar para la fase de aprendizaje y el subconjunto restante para testeo.

- Elección de puntos al azar? Puntos representativos?

Tabla de contenidos

Eta dinámico (η adaptativo)

Incrementar η si el error sube consistenemente incrementar eta aritméticamente: quizás se está siendo muy conservador.
Si el error incrementa, reducir η exponencialmente.

Momentum

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

Cambios dependen de cambios anteriores. Idea de dirección general del error.
Olvido exponencial de cambios anteriores.
Se aplica a cada batch / lote

Ruido

Idea: Escape del mínimo local.
Robustez de la red neuronal: debería poder soportar ruido.

Tabla de contenidos

Resultados 1

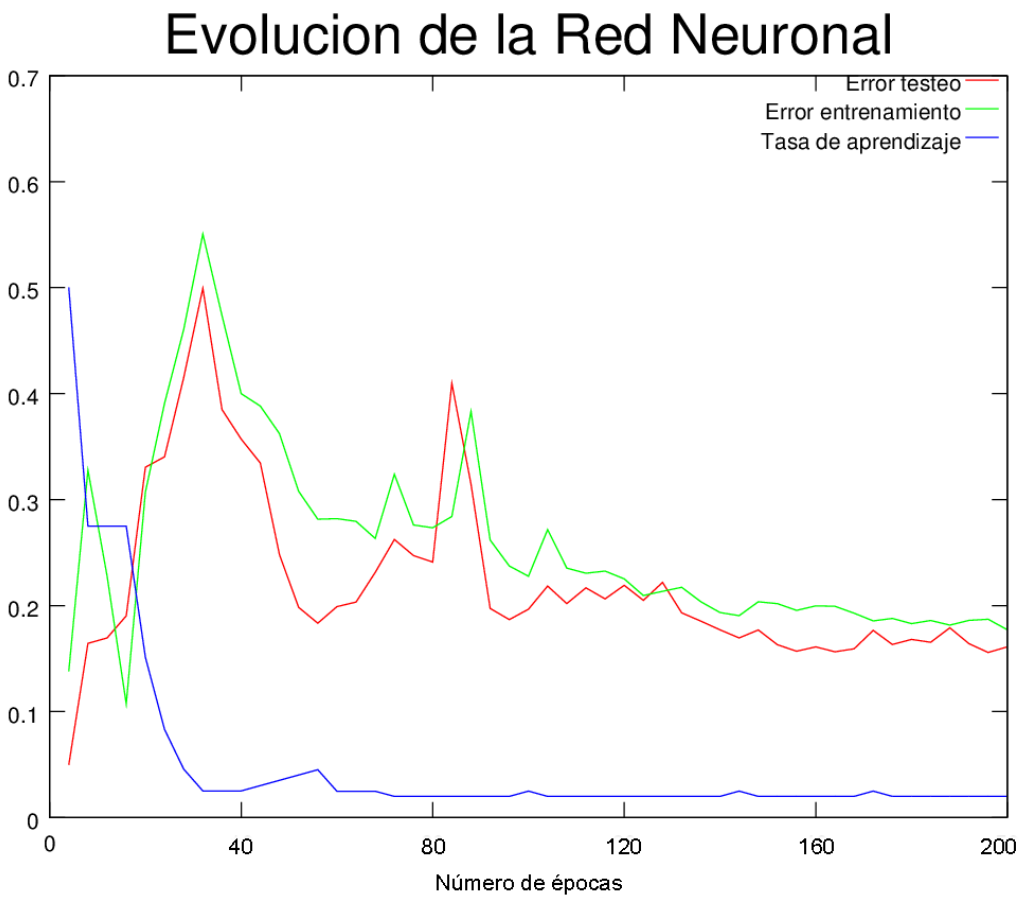

Figura 1: Arq [200 100], eta adaptativo, tangente hiperbólica

Resultados 2

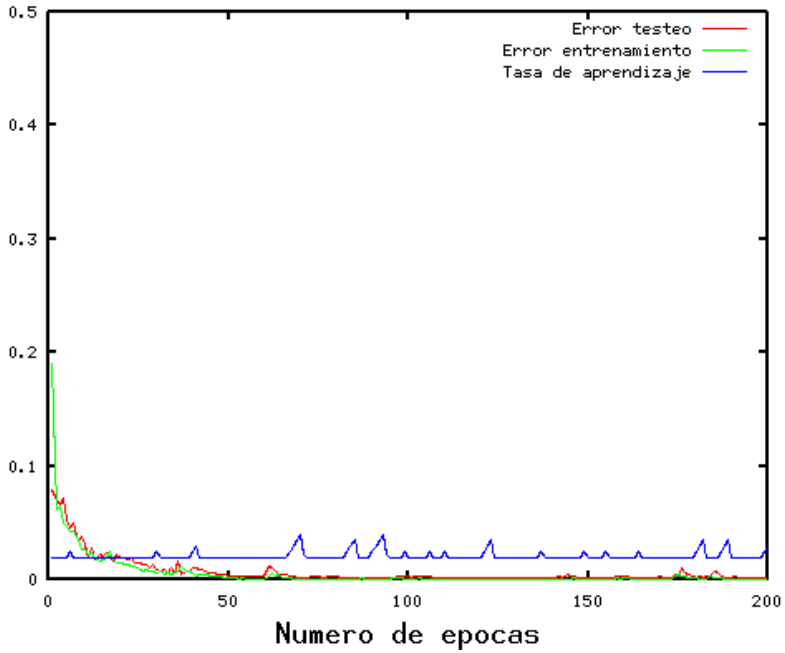

Figura 2: Arq [50 30] 200 epocas, tangente hiperbólica.

Resultados 3

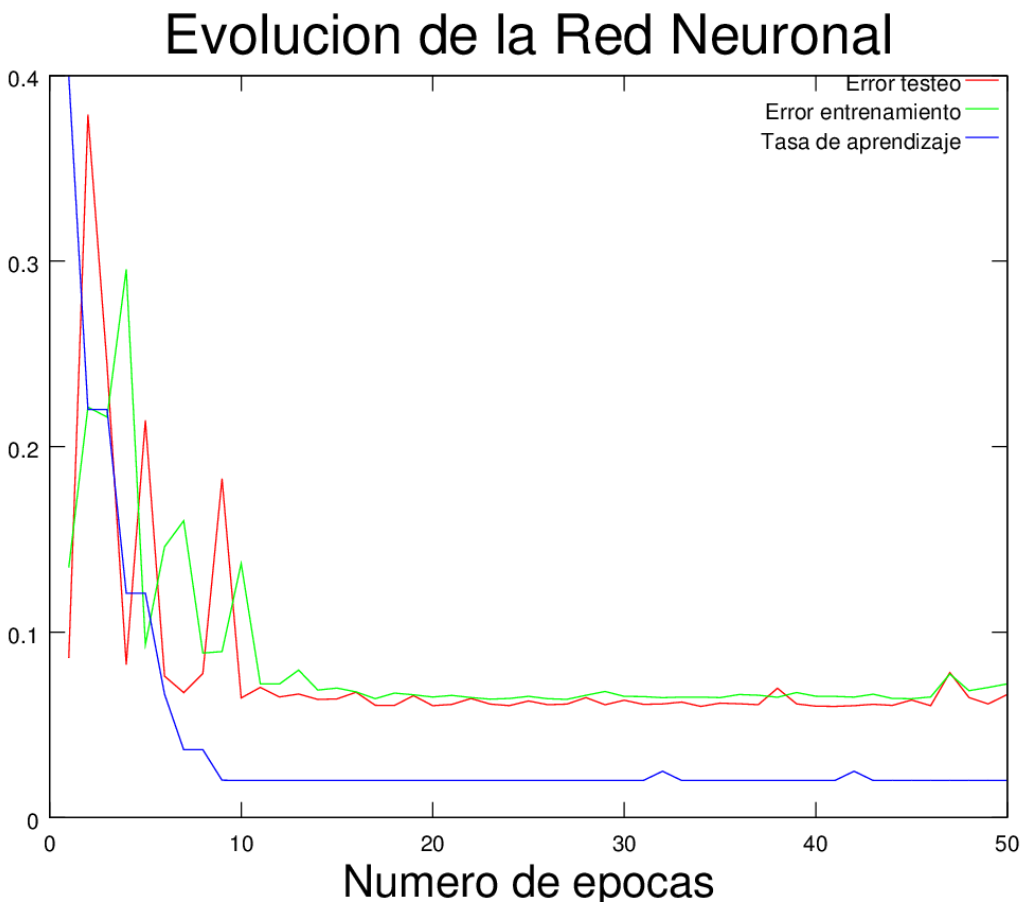

Figura 3: Arq [4 4] 200 épocas eta adaptativo, tanh.

Tabla de contenidos

Conclusión

- Para la resolucion de problemas no siempre es necesaria una mayor cantidad de neuronas para lograr una precisión aceptable (puede llevar a malas generalizaciones y tiempo de más hasta alcanzar el error desdeado). Esto puede observarse en la figura 1 del **Anexo A**
- El aprendizaje de las redes y la velocidad de convergencia dependen exclusivamente de la naturaleza del problema y de los parámetros adecuados para dicho problema en particular.
- Las mejoras al algoritmo de back propagation, no garantizan mejoras en la velocidad de convergencia para todos los problemas.
- Momentum no siempre puede acelerar la convergencia.
- La función de activación *exponencial* es más propensa a atascarse en mínimos.

Conclusión cont.

- Tomar demasiados patrones de entrenamiento pueden perjudicar la capacidad de la red de generalizar, sin embargo muy pocos pueden provocarlo también (muestra no representativa).
- Momentum no siempre puede acelerar la convergencia.
- La función de activación *exponencial* es más propensa a atascarse en mínimos.
- Tomar demasiados patrones de entrenamiento pueden perjudicar la capacidad de la red de generalizar, sin embargo muy pocos pueden provocarlo también (muestra no representativa).