



Universidad Simón Bolívar
Departamento de Computación y Tecnología de la Información
CO-6612, Introducción a las redes neuronales
Trimestre Enero-Marzo 2023
Tarea 4

Eros Cedeño 16-10216

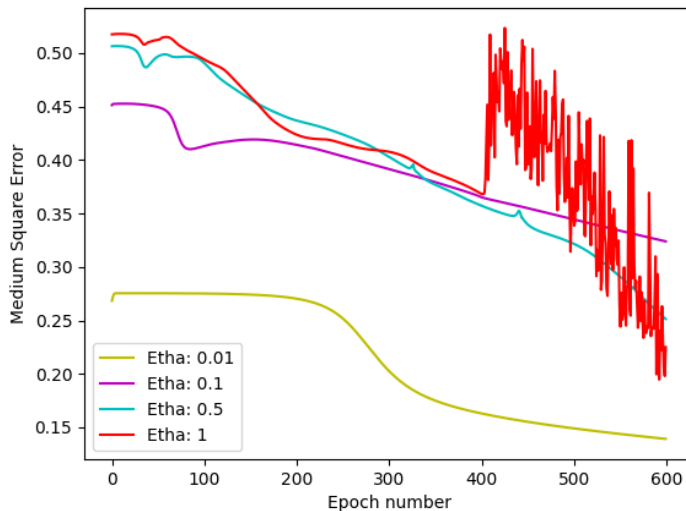
Perceptrón Multicapa MLP

Para las entregas anteriores se han creado dispositivos basados en redes neuronales artificiales como el perceptrón multi clase y el adaline mediante el algoritmo LMS, sin embargo el campo de la inteligencia artificial ha avanzado aún más creando dispositivos más potentes aunque más complejos. Este es el caso del perceptrón multicapa o abreviado el MLP (Multi Layer Perceptron). Este dispositivo es por excelencia el ejemplo y referente de las redes neuronales. Su premisa principal es que mediante la interconexión de capas de neuronas es posible generalizar de más efectivamente dependiendo de las necesidades del usuario. Para esta tarea se realizó la implementación del Perceptrón Multicapa (MLP) en el lenguaje de programación Python 3. Se creó además un programa cliente que entrena dos redes neuronales para clasificar entre dos categorías de datos cada una. La primera de ellas clasifica entre datos de Ciencias de la Tierra y el espacio VS Ciencias Médicas, y la segunda clasifica entre Ciencias de la vida VS Agricultura.

Inicialmente se creó una estructura PerceptronLayer que en resumidas cuentas es una capa de la red del MLP. Es parametrizable la tasa de aprendizaje, las dimensiones de la capa, las dimensiones de cada entrada, el sesgo entre otros. La capa además tiene la capacidad de ser activada dado un estímulo y entrenada dando un estímulo y la capa siguiente si es una capa oculta o la respuesta esperada si es la capa de salida. Luego se creó la clase MLP que crea y enlaza múltiples capas según las dimensiones y parámetros indicados. Y finalmente se elaboró el programa cliente que realiza las pruebas y genera las gráficas.

Cada dispositivo se entrenó durante 600 épocas máximas y se varió la tasa de aprendizaje entre los valores 0.01, 0.1, 0.5 y 1.

En comparación con la tarea anterior, al igual que en el adaline, el MLP busca conseguir un hiperplano tal que minimice lo más posible el error de la clasificación. En este caso la medida para estudiar la efectividad del MLP no fue únicamente el error cuadrático medio, sino que se estudió su capacidad de generalizar. Para esto se separaron los datos en dos conjuntos: Los datos de entrenamiento con aproximadamente el 80% de los estímulos, y los datos de validación con el resto de los estímulos.

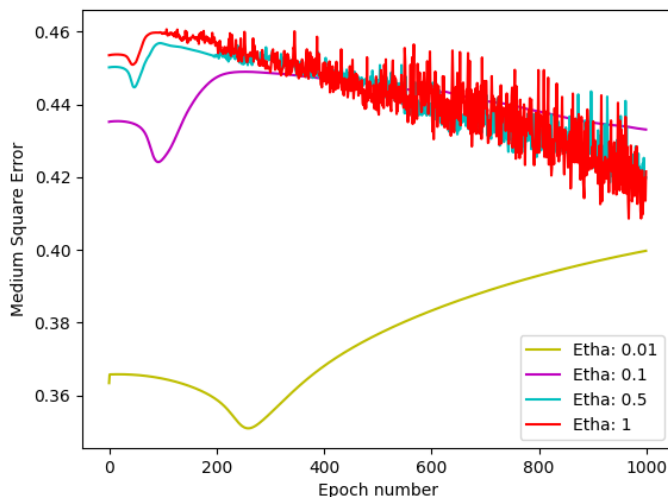


Para la primera clasificación se obtuvo que la tasa de aprendizaje $\eta = 0.01$ tuvo el mejor desempeño al momento de minimizar el error durante las 600 épocas, tal y como se puede observar en la gráfica de la izquierda. Luego se procedió a realizar la validación de la capacidad de la red para generalizar los resultados, para

ello se intentó predecir los datos del conjunto de validación y se obtuvieron resultados adecuados con un error de $1E-05$ en promedio por cada estímulo lo que es impresionante considerando que los datos a comparar no son ni por asomo linealmente separables y en las entregas anteriores como en el perceptrón por ejemplo la tasa de correctamente clasificados fue en general muy mala.

En el perceptrón la convergencia estaba garantizada para cualquier tasa de aprendizaje η por grande o pequeña que sea siempre y cuando los datos sean linealmente separables por otro lado, en el caso del Adaline no era necesaria la separabilidad lineal sino que la escogencia de la tasa de aprendizaje era el factor determinante en el entrenamiento. Ahora en este caso con el MLP tampoco es necesaria la separabilidad lineal y como en el adaline la escogencia de la tasa de aprendizaje es decisiva para lograr minimizar el error cuadrático medio pero para el MLP hay otro factor decisivo: La Topología de la Red. En esta tarea se experimentó con MLPs de 2×1 y 315×1 pero en el caso de la primera no se obtuvieron buenos resultados al generalizar. Y en la segunda el hardware era sumamente ineficiente para entrenar la red por lo que no se pudo entrenar por más de 30 épocas y con resultados de generalización aún muy poco refinados. Finalmente se optó por una red de la forma 4×2 donde el vector de salida tiene dos dimensiones y tiende a 1 en la coordenada

asociada a la clase que corresponde y a 0 en las que no. Para comparar el impacto de la tasa de aprendizaje se ejecutó el algoritmo sobre los casos a estudiar y se observa como algunas tasas son mejores para minimizar el error cuadrático. Pero en contraste, se ejecutó el mismo algoritmo sobre los mismos datos pero fijando la tasa de aprendizaje en 1 se observa como la curva (marcada en rojo) es muy inestable y no tiende monótonamente al mínimo.



Para el segundo caso, al clasificar entre ciencias de la vida y agricultura la tasa sobre topología indicada anteriormente con 1000 épocas se concluye que la tasa con mejor rendimiento al entrenar es $\eta = 0.01$ y entrenando hasta la época 200 aproximadamente. Al validar, el error promedio por clasificación es de aproximadamente $1E-01$.

Adicionalmente quiero agregar que para estos experimentos se utilizó la función Sigmoide como función de activación de todas las capas de la red MLP. Sería interesante que para futuras implementaciones se utilice una función lineal para activar la última capa o capa de salida esto permitirá ajustar y corregir esos pequeños márgenes de error en los resultados cuando se debe generalizar, sin embargo por simplicidad de la implementación se mantuvo la misma función para la totalidad de la red. Se sugiere utilizar la función ReLu con los ajustes necesarios para la capa de salida por su naturaleza rectificadora.

Luego intentando mejorar el rendimiento se agregó el parámetro del momentum y para el primer caso de ciencias de la tierra vs ciencias médicas se logró una mejora en la convergencia con la tasa de aprendizaje 0.01 y el parámetro alpha de 0.3. El gráfico muestra una convergencia más veloz al mínimo global

