



**UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE INGENIERIA
DEPARTAMENTO DE ELECTRONICA
FUNDAMENTOS DE INTELIGENCIA COMPUTACIONAL
TRABAJO 2**

Fecha de entrega: abril 10 de 2018 en clase (No se tendrá en cuenta las tareas enviadas por correo).

METODOLOGÍA

El siguiente trabajo debe ser resuelto en los grupos de trabajo (2 personas). Los resultados se deben entregar en formato de artículo IEEE (un ejemplo lo encuentran en: http://www.ieee.org/conferences_events/conferences/publishing/templates.html) y el trabajo debe ser impreso en clase. Este documento no debe pasar de 5 páginas.

La entrega debe ser puntual, si no se entrega a tiempo el trabajo será calificado sobre 4 y si se entrega después del 12 de abril se calificará sobre 3.

Los problemas los puede resolver empleando cualquier software (ejemplo matlab, python). Vía correo electrónico, se debe enviar el código que se haya empleado o los archivos generados. **No se pueden utilizar funciones de algún toolbox que implementen el entrenamiento de redes neuronales. El envío de este correo electrónico debe hacerse antes de entregar el trabajo impreso.**

EVALUACION

El 50% de la nota corresponde al informe que entregan por escrito, el otro 50% se califica de manera individual y corresponde a la sustentación del trabajo. El punto 1 tiene un valor del 15%, el 2 del 15% y el 3 del 20%, sumando así el 50% correspondiente al informe escrito.

OBJETIVOS

- Afianzar los conceptos de la teoría de redes a través de la implementación de los algoritmos de entrenamiento de RNA.
- Aplicar los conceptos de redes neuronales para la solución de un caso particular de clasificación.

1. Descargue el conjunto de datos sobre la eficiencia en el consumo de energía para la refrigeración y la calefacción de edificios con diferentes formas, que encontrarán en el Machine Learning Repository:
<http://archive.ics.uci.edu/ml/datasets/Energy+efficiency>.

El objetivo de este problema es desarrollar un sistema que permita hacer la predicción de la variable Y1 (Y2 no se analizará). El conjunto de datos comprende 768 muestras y 8 características.

- Divida el conjunto de datos de manera que pueda realizar una validación cruzada o bootstrapping.
- Entrene una neurona usando el subconjunto de entrenamiento utilizando el algoritmo de aprendizaje **LMS** y el algoritmo **Regresión Lineal**. Trate de obtener el menor error de clasificación en el conjunto de prueba que pueda. Para esto intente diferentes tasas de aprendizaje y pesos iniciales, recuerde tener en cuenta la generalización. Compare los dos algoritmos.
- Calcule el error de clasificación en el subconjunto de prueba.
- Analice los resultados obtenidos.

2. Utilice el conjunto de datos “letter recognition” del repositorio de bases de datos de la Universidad de California en Irving (adjunto, formato excel). El vector de entrada consiste en una serie de atributos de una imagen en blanco y negro, que debe ser clasificada como una de las 26 letras del alfabeto inglés.

Para convertir el problema a clasificación binaria, divida las 26 letras en dos clases de la siguiente forma:

- clase 1: letras presentes en el primer apellido de cada uno de los integrantes del grupo
- clase 2: el resto de las letras

La primera columna de cada fila corresponde a la letra que es representada por el vector de la fila.

- Utilice una red neuronal de una neurona para resolver este problema. Para el entrenamiento debe usar el algoritmo de aprendizaje de **perceptrón** y el algoritmo **regresión logística**. Para esto tenga en cuenta:
 - Los parámetros de entrenamiento.
 - Divida el conjunto de datos de manera que pueda realizar una validación cruzada o bootstrapping.
 - Experimente con parámetros hasta encontrar una red que según su criterio sea lo mejor posible. Justifique esta decisión.
 - Compare el rendimiento y el tiempo de entrenamiento de los dos algoritmos.
 - Debe mostrar que la red tiene una probabilidad de error pequeña, recuerde que el objetivo no es tener una red que responda bien a unos datos sino a nuevos datos que reciba. Tenga en cuenta el concepto de Generalización.
3. El archivo “DataNN.txt” contiene una base de datos de cantos de especies de animales (anuros y aves). Cada muestra de este archivo corresponde a una vocalización, a la cual se le extrajeron 76 descriptores (en el archivo de audio se tiene el nombre de cada característica). Se desea obtener un clasificador que permita reconocer cuando entre una nueva muestra, a cuál de las 12 clases (12 especies de animales) esta muestra pertenece.
- Entrene una red neuronal tipo perceptron multicapa utilizando el algoritmo de aprendizaje de **Back propagation**. Para esto tenga en cuenta:
 - La arquitectura de la red, la división de los datos en entrenamiento y prueba.
 - Los parámetros de entrenamiento
 - Divida el conjunto de datos de manera que pueda realizar una validación cruzada o bootstrapping.
 - Experimente con diferentes redes y parámetros hasta encontrar una red que según su criterio sea lo mejor posible. Justifique esta decisión
 - Debe mostrar que la red tiene una probabilidad de error pequeña, recuerde que el objetivo no es tener una red que responda bien a unos datos sino a nuevos datos que reciba. Tenga en cuenta el concepto de Generalización.
 - No es necesario que use las 76 características, es muy común que unas pocas representen la mayoría de la información. Realizar alguna selección de características, se deja como bonus de una unidad en la nota del informe o de la sustentación. La solución no debe ser necesariamente algún algoritmo propuesto, sino algo que quizás a usted se le ocurrió y probó.

Nota: Los algoritmos de entrenamiento debe programarlos usted mismo, no se aceptan las rutinas de entrenamiento de RNA que ya existen en diferentes programas.