UNIVERSIDAD SIMON BOLIVAR
Departamento de Cómputo Científico
CO-6612, Introducción a las redes neuronales
Tarea 4: Perceptrón Multicapas

1. Los datos en *reglin* corresponden a un problema de regresión lineal. Los datos contenidos en dichos archivos contienen un conjunto de 60 puntos para el entrenamiento, 10000 puntos para datos de prueba y los datos que corresponden al modelo real.

   (a) Entrene una red neuronal con $n = [1, 2, 3, 4, 6, 8, 12, 20, 40]$ neuronas en la capa oculta y una neurona lineal en la capa de salida. Utilice el algoritmo de descenso de gradiente por 700 épocas. (no utilice paquetes)

   (b) Grafique el error cuadrático medio del entrenamiento y del conjunto de prueba en cada caso. Comente sus resultados. ¿Cuál es el mejor valor para n?

   (c) Grafique el error del conjunto de prueba así como las funciones entrenadas (relación entrada-salida) para $n = 2, n = 8$ y $n = 40$. Interprete sus resultados. ¿Es decreciente el error sobre el entrenamiento y la prueba?

   (d) Emplee una aproximación polinómica a este conjunto de datos mediante un Adaline. Cuál es el grado del polinomio que mejor interpola estos datos?

2. Los datos *rabbit* muestra la dependencia del peso de la retinas de ciertos conejos australianos en función de su edad. Se tiene que un modelo no lineal encontrado usando mínimos cuadrados sobre los datos, está descrita por la ecuación:

   $$Y = 233.846(1 - \exp(-0.00604x)) + error$$

   Diseñe e implemente su propio perceptrón multicapas que aproxime los datos (no utilice paquetes). Justifique su respuesta (criterio de parada, arquitectura de red, tasas de aprendizaje, algoritmo de aprendizaje, etc.) y aproximación usando todos los recursos disponibles para argumentar su respuesta.

3. Sean $x \in R^d$ los patrones, $w \in R^d$ el vector de pesos, t la respuesta deseada, $\lambda > 0$ una constante de regularización y m el número total de datos. Sea $R(w)$ la función de costos regularizada.

   $$R(w) := \frac{1}{m}\sum_{i=1}^{m} l(t_i - < w, x_i >) + \frac{\lambda}{2}||w||^2, \text{ con } l(\xi) = \begin{cases} \frac{1}{2}\xi^2 & |\xi| < 1 \\ |\xi| - 1/2 & \text{caso contrario} \end{cases}$$

   (a) Calcule el gradiente (en lote) de $R(w)$ con respecto a w.

   (b) Escriba el algoritmo de descenso de gradiente *estocástico* para minimizar $R(w)$.