

OPENCOURSEWARE
REDES DE NEURONAS ARTIFICIALES
Javier Huertas Tato



Enunciado Caso Práctico III: PM (Clasificación)

1. Resumen

En esta práctica se aprenderá el uso del Perceptrón Multicapa (PM) para tratar un problema de clasificación. Los problemas de clasificación y regresión requieren métodos distintos y es necesario tratar los datos de acuerdo a las necesidades propias de dicho problema.

En este enunciado se pide resolver un nuevo problema de clasificación usando el PM. Se completará una experimentación haciendo uso de un script y ajustando los parámetros de dicho script. Debido a las diferencias con regresión no es posible comparar con anteriores métodos, sin embargo, la clasificación requiere un análisis de los resultados diferente y, en este problema, la una buena interpretación de dichos resultados es imprescindible en una buena experimentación.

Al igual que en prácticas anteriores, también es necesario preprocesar los datos (revisar documento [PR-F-01]).

2. Materiales

Se necesitará:

- Dominio de datos de Jaen [PR-F-03.2] (Leer su descripción *Dominio 2 – Nubes Jaen* [PR-F-03.1])
- Script *PM_Clas.R* [PR-F-06.2]
- Intérprete del lenguaje R
- Entorno de programación RStudio

3. Guía para la elaboración

Los pasos a seguir para completar la práctica son:

- Preprocesado de datos:
 - Normalizado de los datos entre 0 y 1
 - Aleatorización de la ordenación de los ejemplos de entrada
 - Separación de los datos en entrenamiento, validación y test (60%, 20% y 20% respectivamente). Dado que este problema es de clasificación, es necesario que haya la misma proporción de clases en distintos conjuntos.
 - Mantener el formato de salida como csv, separando los datos con “;” y los decimales con “,”. Hay que mantener la cabecera con los nombres de las variables.
 - Los archivos que contengan los datos deben tener los siguientes nombres:

- *DatosTrain.csv*
 - *DatosValid.csv*
 - *DatosTest.csv*
- Uso del script de PM:
 - El Script contiene los parámetros en 3 variables distintas: *topologia*, *razonAprendizaje*, *ciclosMaximos*.
 - La topología se expresa como un vector de número de neuronas, así una red de una capa y cinco neuronas se representa como $c(5)$, una red de dos capas, una de cinco y otra de diez, se representa como $c(5,10)$.
 - La razón de aprendizaje y los ciclos de aprendizaje se determinan con un número.
 - Mover el *workspace* de RStudio a la carpeta que contenga el fichero *PM.R*
 - Para lanzar el script desde RStudio tan solo es necesario hacer `source("PM.R")`
 - Tras ejecutar el script se guardarán los resultados en los archivos indicados, además se pueden consultar los resultados escribiendo por consola el nombre de la variable que los contiene.
- Experimentación con PM
 - El perceptrón multicapa tiene varios parámetros que ajustar. La topología, la razón de aprendizaje y el número de ciclos.
 - Los requisitos mínimos de una experimentación son:
 - Probar 3 topologías de una capa y 1 topología de varias capas
 - Probar 3 razones de aprendizaje
 - Ajustar el número de ciclos se puede hacer:
 - Estableciendo un número arbitrariamente alto de ciclos
 - Reducir el número de ciclos hasta que no se produzca sobreaprendizaje (esto es mucho más sencillo revisando las gráficas que muestra el script)
 - Al final debería haber como mínimo 12 experimentos completos con la precisión y precisión por clase de resultado.
 - Cada experimento debe tener documentado su precisión y precisiones por clase.
 - El experimento final debe tener sus matrices de confusión, gráfica de evolución del error, salidas y red guardados.
- Extraer conclusiones
- Memoria
 - Realizar una memoria breve que muestre la elaboración de una experimentación completa y los pasos realizados durante la práctica.