

Master Universitario. Análisis de Datos Masivos (MADM)
Assignatura: 11644 - Aprendizaje Estadístico y Toma de Decisiones II
Elemento de Evaluación: Prueba de mínimos on-line
Cuestiones sobre Redes Neuronales Artificiales

Fecha de entrega: 30 Mayo 2021.

Se considera la base de datos “bank_full.csv”. Los datos están relacionados con una campaña de marketing directo de una institución bancaria portuguesa. Las campañas de marketing se basaban en llamadas telefónicas. A menudo, se requería más de un contacto con el mismo cliente, para saber si el producto (depósito bancario a plazo) sería (“sí”) o no (“no”) suscrito. La descripción de la base de datos se puede encontrar en el fichero “bank_names.txt”. Observar bien la descripción de los datos ya que pueden existir variables que no son útiles para nuestro estudio. Se trata de resolver un problema de clasificación usando ANN sobre el fichero “bank_full” y siguiendo los pasos descritos en las clases. Debéis desarrollar un estudio que incluya los siguientes puntos:

1. Exploración de datos: se trata de estudiar los datos en R. Conviene estudiar la existencia de “missing values” e identificar datos atípicos.
2. Visualización de datos: se trata de representar gráficamente los datos. Si usáis boxplots para datos numéricos podéis ver mejor los outliers y decidir si los elimináis o no. Para las variables categóricas utilizar histogramas.
3. Partición de los datos 70:30.
4. Ajuste del modelo y plots. Utilizar una red neuronal artificial para clasificar los datos. Hacer diversas pruebas cambiando la estructura de la red y los parámetros. Usar cross-validation. Proporciona una comparativa de las redes utilizadas.
5. Matriz de confusión. Es importante que comentéis los resultados que vayáis obteniendo.

Observaciones:

Se tiene que entregar una memoria de la práctica (en formato html y Rmd). Así como todos aquellos ficheros que sean necesarios para obtener los resultados que mostréis en el informe de la práctica.
