

Informe Assignment 1

ICAI. Machine Learning.

Álvaro Rodríguez González, Pablo Sanz Caperote

Curso 2021-22. Última actualización: 2021-10-24

Índice

Análisis exploratorio de los datos.	3
Regresión Logística	4
Caso general	4
Optimizada	4
KNN	4
Árboles de decisión	4
Caso general	4
Optimizada	4
SVM	4
SVM Lineal	4
SVM Radial	4
Redes Neuronales	4
Caso general	4
Optimizada	4
Random Forest	4
Comparación de modelos	4
Conclusiones	4

Análisis exploratorio de los datos.

Comenzamos cargando tanto los datos como las librerías que necesitaremos para el desarrollo de nuestros modelos. Tras ello hacemos una primera exploración rápida de estos a través de la tabla de R.

A continuación lo primero que hacemos es ver como está estructurado nuestro conjunto de datos y nos damos cuenta de que todas las variables son de tipo numérico (int o num). Esto nos supone un problema para trabajar con los modelos ya que necesitamos que nuestra variable de salida (en este caso será DIABETES) sea un factor. Por ello transformamos la variable DIABETES en factor, donde también cambiamos los 0's por "No" y los 1's por "Si" ya que si dejásemos los 0's y 1's no estaríamos trabajando correctamente con factores.

Tras este cambio lo que haremos será hacer un summary de la tabla de datos con la finalidad de ver si esta tiene algún valor nulo (NA's). La función nos devuelve que no existe ningún NA, por ello podemos pasar a buscar valores atípicos dentro de nuestras variables.

Para encontrar los outliers lo primero que haremos será una representación gráfica de todas las variables juntas (un ggpairs) con la finalidad de ver como se comportan.

Regresión Logística

Caso general

Optimizada

KNN

Árboles de decisión

Caso general

Optimizada

SVM

SVM Lineal

SVM Radial

Redes Neuronales

Caso general

Optimizada

Random Forest

Comparación de modelos

Conclusiones