Estimados alumnos.

Tal como se señaló en el anterior mensaje se ha procedido a subir el resto de documentación necesaria para la primera tarea. Cada una de las partes o temas, para facilitar el archivo, se ha estructurado en cuatro carpetas: Teoría, Ejemplos, Vídeos, Material complementario.

Al margen de esta documentación, un sitio muy recomendable de consulta es CRAN (Comprehensive R Archive Network), [https://cran.r-project.org](https://cran.r-project.org/) donde se dispone de numerosos manuales y artículos de investigación, con ejemplos ilustrativos de programación de las diferentes técnicas.

Dejo a continuación la primera tarea a realizar, la cual se estructura, lógicamente, en dos partes, una correspondiente al uso general de R y otra donde hay que hacer un preprocesado de la información y un descriptivo.

Notas:

* El descriptivo se hará, lógicamente, antes de hacer el equilibrado, pero después de tratar con los posibles outliers y con los posibles datos perdidos o ausentes.
* Para aquellos que quieran presentar resultados de modelar, la regresión logística se hace con la función glm, es decir:

modelo <- glm(CLASE ~ ., data = datos.ok, family = "binomial")

o pueden usar otros tipos de modelos como random forest (librería y función randomForest o k-vecinos (librería class y función knn), consulten siempre el help. Por ejemplo, si queremos ayuda de la función knn (tenemos que cargar antes la librería class) tecleamos ??knn, y se tiene una descripción y ejemplos ilustrativos a mano. Además de, como se ha dicho, CRAN, otro sitio interesante es RPubs, donde investigadores cuelgan mucho código. Investiguen en web y compartan cosas con sus compañeros en el foro. R tiene muchas librerías.....

* Dicho esto, el objetivo es plantear cuáles pueden ser las variables más influyentes o más relacionadas con la variable CLASE, y no tanto hacer un ajuste fino de un modelo.

Atentamente,

el equipo docent