

# Laboratorio de programación en R

## Minería de datos: Aspectos Avanzados

7 de Febrero, 2018

Para esta sesión de laboratorio, se necesitará descargar los ficheros `ordinal.Rmd` y `esl-j48ordinal.R` desde PRADO, además del conjunto de datos `esl.arff`.

Para esta evaluación **opcional**, tendréis que subir dos ficheros R `esl-j48ordinal.R` y `generalordinal.R` a la actividad correspondiente de PRADO, antes del **22 de Febrero de 2018**. Hay que asegurarse de que el código entregado esté suficientemente comentado. También, proporcione un análisis breve de los resultados obtenidos usando comentarios en los ficheros R.

### Clasificación ordinal

En este ejercicio, se implementará la técnica de múltiples modelos para clasificación ordinal utilizando el algoritmo J48 del paquete RWeka de R. Para ello, disponemos de una plantilla `ordinal.Rmd` (`ejemploOrdinal.pdf`) que hace el proceso de múltiples modelos considerando el clásico data set iris y el clasificador J48 de Rweka. En este caso, se ha supuesto que las clases de iris tienen una disposición ordinal. Como iris tiene tres clases, se generan dos conjuntos de datos desde el original. Después, se contruyen dos árboles asociados a cada conjunto. Finalmente, éstos se utilizan para clasificar una instancia de test seleccionada y obtener las probabilidades.

### Modelos múltiples para el data set `esl.arff`

Se pide modificar el fichero `esl-j48ordinal.R` para adaptar el proceso de modelos múltiples de clasificación ordinal al conjunto de datos `esl` con el algoritmo J48 (C4.5). Para ello, se considerará una única partición entrenamiento test aleatoria con 100 ejemplos en el conjunto de test. Este conjunto tiene carácter ordinal y dispone de 9 clases [1, 9]. Una vez generados los modelos desde el conjunto de entrenamiento, habrá que clasificar los 100 ejemplos de test usando la cascada de probabilidades que define el modelo múltiple. El resultado se generará en un fichero llamado `esl-j48ordinal.R`.

### Generalización del proceso de modelos múltiples

El objetivo de esta segunda parte es la de construir este proceso de forma genérica para cualquier data set ordinal y un clasificador diferente al J48. Se implementará una función que recibirá un data set como parámetro y será capaz de formar los data sets binarios derivados a partir del número de clases del data set original. Así mismo, lanzará un clasificador diferente al J48 (a elección por el estudiante) para cada uno y devolverá una colección de modelos construidos.

Una segunda función recibirá dicha colección y será capaz de predecir un ejemplo (o un conjunto de ejemplos) a partir de los mismos usando la cascada de probabilidades que define el modelo múltiple. El resultado se generará en un fichero llamado `generalordinal.R`.