

# M2.2.2 Modelos Supervisados y No Supervisados

## Programa Big Data y Business Intelligence

Enrique Onieva

[enrique.onieva@deusto.es](mailto:enrique.onieva@deusto.es)

<https://twitter.com/EnriqueOnieva>

<https://www.linkedin.com/in/enriqueonieva/>

# Nota

- A continuación se presentan una serie de ejercicios sugeridos para la mejor comprensión de los conceptos vistos en clase, así como para el desarrollo práctico con las herramientas trabajadas.
- Postea avances y dudas surgidas en el foro, para un aprendizaje conjunto a tus compañeros.

# KNN

- Ejecute KNN sobre el dataset dentro del fichero "car.csv".
  - Como el dataset no es numérico, haga dos pruebas:
    - Convirtiendo todas sus variables en dummy (función `model.matrix`)
    - Convirtiendo los niveles en numéricos (Low=1, Med=2, High=3,...)...
  - Haga validación cruzada para obtener el mejor valor de K en cada caso
- ¿Cómo obtenemos mejores resultados? ¿Por qué crees que ocurre ésto?

# Redes Neuronales (I)

- En teoría, el número de variables "inútiles", no debe de afectar al rendimiento de la red, pero sí al tiempo necesario para aprender
  - ¿Puedes ponerlo en práctica como hemos hecho para otros métodos?
    - Crear 1, 2, 3, ... columnas aleatorias a un dataset, para ver cómo afecta a la calidad de los resultados obtenidos por la red
    - Medir también el tiempo necesario para su aprendizaje según el número de columnas añadidas
      - Para medir tiempos puede usar la función [proc.time\(\)](#)

# Redes Neuronales (II)

- Las redes neuronales requieren de un arduo proceso de ajuste de parámetros, así como de prueba de diferentes estructuras neuronales.
  - ¿Puedes ajustar una red para obtener los mejores resultados posibles para los problemas de clasificación (Churn) y regresión (hormigón)?
  - Recuerda el hacer uso de validación cruzada para garantizar fiabilidad en las comparaciones que hagas

# Árboles de Decisión

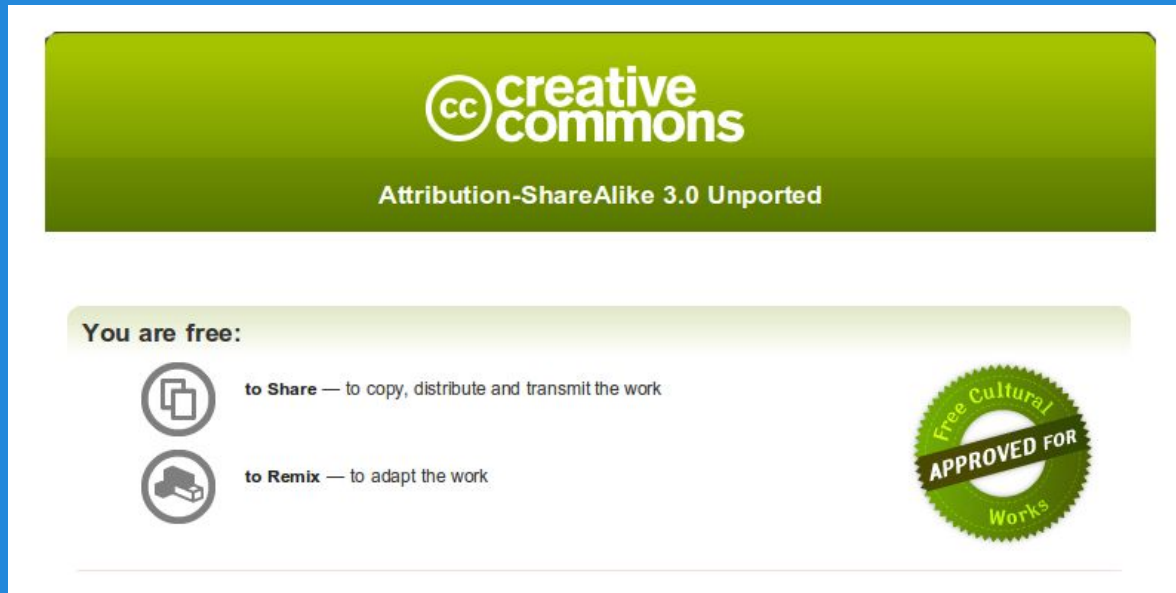
- Los árboles de decisión pueden ser (siempre que su tamaño sea razonable) supervisados por un experto, para ello, necesitamos visualizarlo convenientemente
- Juega con los (muchos) parámetros de las funciones `prp` y `rpart.plot`
  - A ver quién saca el árbol más molón
  - <http://www.milbo.org/rpart-plot/prp.pdf>

# Clustering

- Os he dejado un dataset de terremotos y otro de jugadores de baloncesto
- Realizar un análisis cluster de ellos:
  - ¿Tienen relación los clusters obtenidos con las posiciones de un equipo de baloncesto?
  - ¿Los grupos de terremotos se asocian mejor por su localización geográfica o por sus características (escala, profundidad...)?

Copyright (c) University of Deusto

This work (but the quoted images, whose rights are reserved to their owners\*) is licensed under the Creative Commons "Attribution-ShareAlike" License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/3.0/>



# Enrique Onieva

[enrique.onieva@deusto.es](mailto:enrique.onieva@deusto.es)

<https://twitter.com/EnriqueOnieva>

<https://www.linkedin.com/in/enriqueonieva/>