

Proyecto 3

“Regresión Lineal”

Intelligent Systems

Descripción del problema

● Data

- El fichero **Baseball.csv** contiene información de 337 jugadores de béisbol.
 - Fuente: <http://sci2s.ugr.es/keel/dataset.php?cod=76>
- El fichero tiene una fila por cada jugador, con los siguientes atributos:

- Batting_average	- On-base_percentage	- Runs
- Hits	- Doubles	- Triples
- HomeRuns	- Runs_batted_in	- Walks
- Strike-Outs	- Stolen_bases	- Errors
- Free_agency_eligibility	- Free_agent	
- Arbitration_eligibility	- Arbitration	Salary

Crea una aplicación web con Shiny (1/3)

- **File -> New File -> Shiny web app**
 - Documentación: <http://shiny.rstudio.com/>
- **La aplicación debe permitir lo siguiente:**
 - Subir un fichero
 - Leer el fichero
 - Preguntar qué columna se debe predecir (“target”)
 - (continua....)

Crea una aplicación web con Shiny (2/3)

- **Cuando la columna a predecir esté seleccionada**
 - Realiza un proceso [5-fold cross validation](#)
 - Para cada partición:
 - Divide los ejemplos en training-set y test-set
 - Crea un modelo lineal que permita predecir el target con cada training-set.
 - Haz la predicción usando cada test-set
 - Acumular el error medio de cada partición
 - Calcula el error medio absoluto de las 5 ejecuciones.

Crea una aplicación web con Shiny (3/3)

- **Después de terminar todo el proceso:**
 - Muestra el valor del error medio absoluto obtenido
 - Muestra un gráfico de barras con el error medio de cada una de las 5 ejecuciones.

Ayuda para la implementación en R

- **Para crear el “fold cross validation”**

- `indices <- createMultiFolds(y=data$target, k=partitions, times=times)`
 - `data = data frame de datos // target = columna a aprender`
- Esta sentencia crea “k” vectores de índices para training-data.
 - `training.data <- data[indices[[i]],]`
 - `test.data <- data[-indices[[i]],]` # importante ‘-indices’

- **Para crear el modelo lineal**

- `linear.model -> lm(target~., training-data)`

- **Para realizar predicciones**

- `prediction <- predict(linear.model, test.data)` # vector con resultados

- **Para calcular el error**

- `abs(prediction - test.data$target)` # ‘target’ es la columna a aprender

Pregunta de teoría

- **Explica con tus propias palabras qué información se obtiene tras ejecutar la siguiente secuencia de sentencias en R:**
 - `data = iris[, -5]`
 - `model = lm(Petal.Width ~ ., data)`
 - `summary(model)`

Formato de la entrega

- **Formato**

- ZIP con el código en R. (nombre de grupo)
 - Script en R (con el código de grupo en el nombre)
 - Documento con la respuesta a la pregunta de teoría

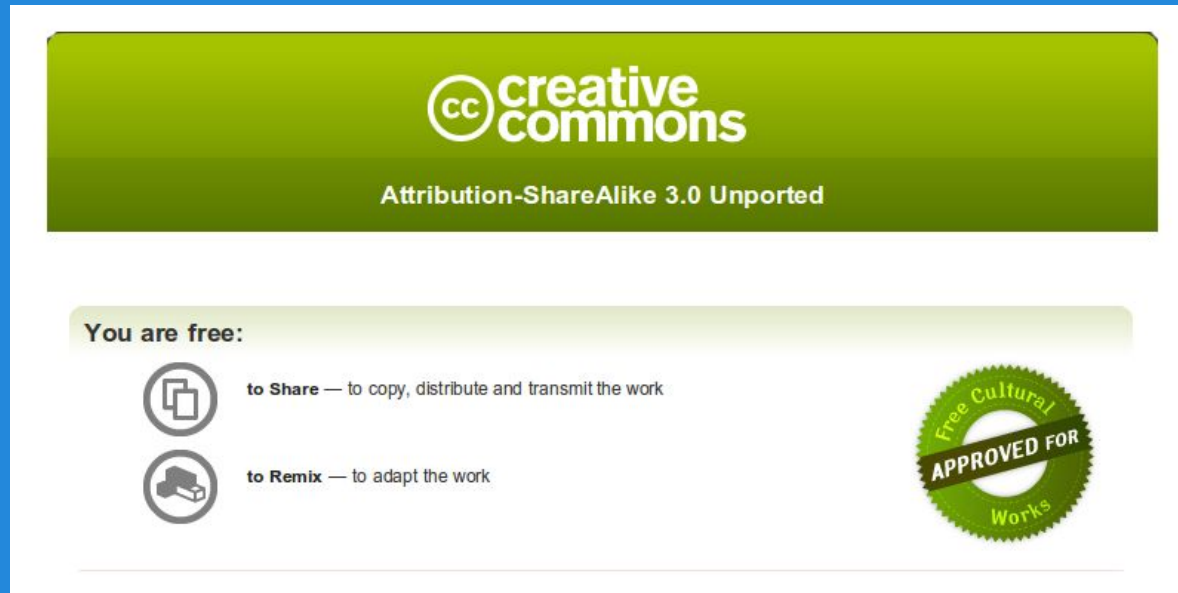
- **Calificación - 10%**

- **Criterios de corrección**

- Corrección de la implementación y la teoría (8,0%).
- Documentación y limpieza del código (2,0%).

Copyright (c) 2016 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/>



Intelligent Systems