



# Week 11

# R

# Day 4

INTRO TO ML

# Caret

<http://topepo.github.io/caret/index.html>



## El paquete **caret** [classification and regression training]

incluye una serie de funciones que facilitan el uso de decenas de métodos complejos de clasificación y regresión.

Otros proyectos similares a caret pero más recientes son: mlr3, tidymodels y H2O



¡Tarda bastante en instalarse! ~ 10 min

```
install.packages("caret", dependencies = c("Depends", "Suggests"))
```

# train()



---

01.

## Formula

Aquí es donde se especifica cual es la variable dependiente (lo que vas a predecir) y la o las independientes (features).

---

02.

## Dataset

Los datos

---

03.

## Método o algoritmo

Se especifica qué modelo de clasificación o regresión vamos a utilizar.

# formula

**[target variable] ~ [predictor variables]**

**# Multiple Linear Regression Example**

**$y \sim x1 + x2 + x3$**

where y is the dependent variable, and x1, x2 and x3 are independent variables. If you want to pass all attributes you can write it as

**$y \sim .$**

```
library(caret)
# Simple linear regression model
model <- train(mpg ~ wt,
               data = mtcars,
               method = "lm")

# Multiple linear regression model
model <- train(mpg ~ .,
               data = mtcars,
               method = "lm")
```

# Functions

## createDataPartition()

Con esta función vamos haremos el train\_test\_split.

```
createDataPartition(  
  y,  
  p = 0.8)
```

## predict()

Para predecir.

```
predict(  
  modelo,  
  newdata = dataset_completo)
```

## postResample()

Esta función nos devuelve las siguientes métricas: RMSE, Rsquared y MAE.

```
postResample(  
  pred = pred_y,  
  obs = real_y)
```

# You got this!



**Relax & keep coding!**

# Ejercicio práctico

En la carpeta data encontrarás un archivo `USA\_housing.csv` y en esta misma carpeta un archivo `.ipynb`.

---

01.

**Teniendo en cuenta estos archivos, realiza una regresión lineal multivariable simulando lo ya realizado en Python.**

Esta vez deberás utilizar lo aprendido con el paquete Caret.

---

02.

**Mejora las gráficas planteadas en clase.**

Usa ggplot2.

---

03.

**Metete en la documentación de Caret y utiliza al menos dos funciones que no se hayan explicado.**

¿Qué hace la función `varImp` y cómo podrías plotearla?

Una vez acabes, sube el archivo a tu proyecto de R en github y haznos llegar el link por email.