# EST-25134: Aprendizaje Estadístico

**Profesor**: Alfredo Garbuno Iñigo — Primavera, 2022.
**Objetivo**. Dar un panorame de lo que rige los principios del curso de aprendizaje estadístico y diferenciarlo de otras estrategias de modelado predictivo. Sentar la notación base.

## 1. INTRODUCCIÓN

Herramientas para `entender reglas de asociación`. Con el objetivo de `generar predicciones acertadas`. No es `estadística`. No es `inferencia causal`.

### 1.1. Caracteristicas del Aprendizaje Estadístico

- Flexibilidad.
- Procesamiento automático.
- Complejidad (en datos).
- Predicción.

En aprendizaje estadístico usualmente consideramos menos suposiciones de los datos en contraste con estadistica; pensamos que el procesamiento es automático; que los datos son complejos; y que el interés primordial es la **predicción**.

### 1.2. Filosofía del curso

Es importante entender los modelos, la intuición y fortalezas y debilidades de los métodos que se utilizan en tareas de aprendizaje `supervisadas` y `no supervisadas`.

### 1.3. Principios

Consideraremos los siguientes, tomados de [1].

- Muchos métodos de aprendizaje son relevantes para una gran variedad de aplicaciones.

No encasillar en el ámbito académico o estadístico. Por supuesto hay muchos mas modelos de los que veremos pero nos concentramos en los que no son de nicho.

- Aprendizaje estadístico **no** es una colección de cajas negras.

Lamentablemente no hay un método que sea exitoso para cualquier tipo de aplicación. Tenemos que conocer bien nuestras herramientas para saber cuál usar en qué situación.

- Una cosa es entender cómo funciona; otra, implementarlo desde cero.

No reinventaremos la rueda. El curso nos concentraremos en las ideas, no en la implementación.

- Conocimiento de dominio. Obtención de información relevante.

> Espero poder transmitir la necesidad de pensar en la aplicación del modelo. Algo que no se ve dentro de una formulación matemática. Pero la cual, si no se es cuidadoso podría tener en consecuencia una herramienta inservible al mediano/largo plazo.

## 2. DISTINCIONES CON RESPECTO A ML

En aprendizaje estadístico nos interesa comprender el proceso que genera los datos y su representatividad estadística.

## 3. MÉTODOS

Aprendizaje `supervisado` y `aprendizaje no supervisado`.

### 3.1. Tareas de predicción

- Clasificación
- Regresión

## 4. NOTACIÓN

Denotamos por $n$ el `número de observaciones`; $p$ para el `número de características` de dichas observaciones. Así que, $x_{ij}$ con $i = 1, \ldots, n$ y $j = 1, \ldots, p$ será un elemento de nuestras observaciones.

Usualmente tendremos una característica que queremos predecir y la denotamos por $y$ . Consideraremos $y \in \mathbb{R}$ ó $y \in \{0, 1\}$ ó $y \in \{1, 2, \ldots, K\}$.

El conjunto de datos que tenemos `disponible para entrenar` modelos lo denotamos por

$$\mathcal{D}_n = \{(x_1, y_1), \ldots (x_n, y_n)\}\,. \tag{1}$$

Es ideal considerar que además tenemos datos adicionales para `hacer pruebas`. A este conjunto lo denotaremos por

$$\mathcal{T}_m = \{(x_1, y_1), \ldots (x_m, y_m)\}\,. \tag{2}$$

Posiblemente necesitemos notación mas especializada para hacer distinciones adicionales o el contexto nos ayude a requerir una notación mas laxa. Esto lo definiremos sobre la marcha.

## 5. AMBIENTE DE R

```
library(tidyverse)    # Herramientas de procesamiento
library(tidymodels)   # Herramientas de modelado
library(ISLR)         # Datos del libro de texto
library(MASS)         # Datos de Boston
```

### 5.1. ¿Por qué utilizamos `tidymodels`?

La búsqueda `CRAN Task View: Machine Learning & Statistical Learning`:

abess (core) ahaz arules BART bartMachine BayesTree BDgraph biglasso bmrm Boruta bst C50 caret CORElearn Cubist deepnet DoubleML e1071 (core) earth effects elasticnet evclass evtree frbs gamboostLSS gbm (core) ggRandomForests glmnet glmpath GMMBoost gradDescent grf grplasso grpreg h2o hda hdi hdm ICEbox ipred islasso joinet kernlab (core)

klaR lars lasso2 LiblineaR maptree mboost (core) mlpack mlr3 mlr3proba mpath naivebayes ncvreg nnet (core) OneR opusminer pamr party partykit pdp penalized penalizedLDA picasso plotmo quantregForest randomForest (core) randomForestSRC ranger rattle Rborist RcppDL rdetools relaxo rgenoud RGF RLT Rmalschains rminer ROCR RoughSets rpart (core) RPMM RSNNS RWeka RXshrink sda SIS splitTools ssgraph stabs SuperLearner svmpath tensorflow tgp torch tree trtf varSelRF wsrf xgboost

## REFERENCIAS

[1] M. Kuhn and K. Johnson. *Applied Predictive Modeling*. Springer New York, New York, NY, 2013. ISBN 978-1-4614-6848-6 978-1-4614-6849-3. . 1