

EST-25134: Aprendizaje Estadístico

Prof. Alfredo Garbuno Iñigo

[Primavera, 2022]

Objetivo

El Aprendizaje Estadístico trata de cómo establecer reglas de asociación al observar un conjunto de datos. En este curso estudiaremos las ideas fundamentales detrás de los técnicas de aprendizaje e incorporaremos con conceptos familiares de estadística. El libro de texto será [?] y lo complementaremos con el enfoque de [?]. Se estudiarán, además, técnicas estadísticas modernas aplicadas a modelado predictivo por medio de lecturas que acompañaran el curso.

El libro de texto será:

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning*. Springer, New York, NY. Second Edition.

Nos apoyaremos de:

- Kuhn, M., Johnson K. (2013). *Applied predictive modeling*. Springer, New York, NY.

Temario

El temario para el semestre de primavera 2022 se puede encontrar aquí. Este es un plan preliminar que está sujeto al avance del curso a lo largo del semestre.

¿De qué se trata el curso?

El material que estudiaremos será sobre modelos predictivos desde el punto de vista estadístico. Dicho de otra forma, no será un curso de *Machine Learning* **algorítmico**. Nos concentraremos en las conexiones entre modelos predictivos y principios estadísticos que le dan sustento.

Recursos

El contenido actual de esta iteración del curso se encuentra en la rama: spring-2022. La estructura del repositorio se muestra a continuación donde tenemos a grandes razgos:

```
.
├── auto
├── css
├── docs
├── images
├── notas
├── renv
└── rscripts
```

7 directories

Horarios

El curso será los días Lunes y Miércoles en un horario de 13:00 a 14:30.

Formato

El curso será presencial. En las primeras dos semanas de clase se tendrá la opción de atender de manera remota en caso de que la situación del estudiante lo amerite. Por otro lado, salvo se indique lo contrario, **las evaluaciones serán presenciales.**

Canvas

Utilizaremos este LMS para gestionar la comunicación fuera de los horarios de clase. Asimismo, ahí tendrán acceso al calendario y las sesiones de zoom que se necesiten a lo largo del semestre.

Fechas importantes

- La sesión del **Lunes 7 de Febrero** será una sesión usual. El miércoles 9 de Febrero **no habrá clase.**
- La sesión del **Lunes 21 de Marzo** hay descanso obligatorio.

- Por solidaridad, el **Miércoles 9 de Marzo no habrá sesión de clase**. Tendremos una sesión de reflexión.

Grabaciones de clases

La clase se enriquece exponencialmente si existe una conversación entre los asistentes. Por lo tanto, las sesiones no se grabarán (sí, incluso si hay un cumpleaños!).

Evaluación

La evaluación del curso será por medio de:

- Participación y tareas (30%)
- Evaluaciones parciales (30%)
- Proyecto final (40%)

La entrega de los trabajos/tareas se realizará por medio de un Github Classroom.

Proyecto final

Tarea de aprendizaje (supervisado/no supervisado) que deseen.

Opciones: Conjunto de datos ó modelo que no hayamos visto en clase.

Entregables: Seguimiento (10%), reporte (20%) y presentación (10%).

Ambiente de trabajo

El curso es agnóstico al lenguaje de programación. Sin embargo, las notas y el material estará construido a partir de R. En particular utilizaremos tidymodels como herramienta de programación dentro de R pues ofrece un ambiente unificado de distintas librerías enfocadas en modelos predictivos. Además, es una herramienta que representa muy bien la filosofía del curso.

Se sugiere utilizar Rstudio para poder trabajar en sus proyectos y sus tareas. En particular, el material de clase será editado en GNU Emacs a través de **orgfiles** (archivos con sufijo **.org**) pero pueden ser visualizados en Github.

Nota: Es necesario tener instalada la versión **4.1.1** de R para tener la mejor compatibilidad con el código del curso.

Configuración R

Utilizaremos **renv** para mantener actualizada las herramientas de **R** junto con el contenido del curso. En la carpeta **notas** se encuentran los archivos de requerimientos (**renv.lock**) con el que podrán usar los archivos que se vayan decantando en la carpeta **rscripts**. Aún así, la configuración se irá construyendo en los ejercicios de tarea que vayamos utilizando en el curso.

Configuración renv

Se recomienda escribir en el archivo `~/.Renviron` la siguiente línea lo cual mantendrá el cache de **renv** en un lugar centralizado

```
RENV_PATHS_ROOT=~/.renv
```

Configuración de Docker

Deseable, y próximo en anunciarse.

Tarea

Las primeras tareas del semestre son básicamente configuración y es recomendable hacerlas si no cuentan con experiencia en: **Git**, **Github**, **R** y el **tidyverse**.