

Introducción al curso

Aprendizaje automático 1

Juan R Gonzalez
juanr.gonzalez@isglobal.org

UAB - Department of Mathematics
BRGE - Bioinformatics Research Group in Epidemiology
ISGlobal - Barcelona Institute for Global Health
<http://brge.isglobal.org>

Temario

- ▶ Introducción a Tidyverse
- ▶ Introducción al aprendizaje automático
- ▶ Regresión lineal y logística
- ▶ Métodos de aprendizaje para datos $n \ll p$
 - ▶ Métodos de penalización (shrinkage)
 - ▶ Métodos de regularización
 - ▶ La librería 'caret'
- ▶ Pasos previos a la creación de un modelo predictivo y medidas de validación
- ▶ Métodos de aprendizaje automático
 - ▶ Árboles de clasificación
 - ▶ K-vecinos más cercanos
 - ▶ Random Forest
 - ▶ Boosting
- ▶ Métodos de aprendizaje para datos grandes (big data)
 - ▶ XGBoost
 - ▶ La librería 'H2O'

Logística del curso

- ▶ Clases lunes de 15:00 a 17:00 (teoría) + 17:00 a 19:00 (prácticas)
- ▶ Clases presenciales aula C5/032 (grupos alternos — alguien sólo online?)
- ▶ Clases online en las horas de teoría, pero puede haber algunas grabadas y luego hacer las 2h de prácticas para no ser tan duro
- ▶ Clases que se hagan en un grupo se grabarán o repetirán otro día
- ▶ Cada semana se pondrá una noticia con lo que se va a hacer cada lunes.
- ▶ Las clases online se harán en Jitsi (<https://jitsi.org/>) (open source) y se comunicará el link en dichas noticias.

Material del curso

- ▶ No seguiremos ningún libro de texto porque estamos tratando muchos temas que están muy bien explicados en varios libros y, sobre todo, en materiales públicos. He creado un bookdown accesible en https://isglobal-brge.github.io/Aprendizaje_Automatico_1/ que describe el contenido del curso
- ▶ Habrán presentaciones en diapositivas que estarán disponibles en el Moodle de la asignatura
- ▶ Habrán link a material opcional
- ▶ Las actividades en el Moodle estarán clasificadas como:
 - ▶ P: puntuable (se evaluará según el contenido)
 - ▶ P2: puntuable (no se evaluará si está bien - se dará la solución: 10 entregado, 5 entregado no completo, 0 no entregado)
 - ▶ I: información
 - ▶ O: opcional

Metodología

- ▶ Clase de teoria: se definen y se explican los diferentes métodos con sus características particulares y se muestran ejemplos concretos.
- ▶ Clase de prácticas: se trabajan los métodos explicados en clase de teoria con diversos conjuntos de datos utilizando el lenguaje de programación R.

Se considera que, para cada hora de teoria y prácticas, el alumno deberá dedicar una hora adicional para la preparación y/o finalización de la sesión.

- ▶ Preguntas de auto-evaluación que se llevarán a cabo en el Moodle que servirá para consolidar los conocimientos aprendidos en las sesiones teóricas.

Evaluación

La nota final se basará en las siguientes notas ponderadas

- ▶ 30% Nota Examen final (tipo test - conceptos)
- ▶ 50% Nota Prácticas (compuesta por prácticas semanales más una práctica final)
- ▶ 20% Nota preguntas de Auto-Evaluación

Se requerirá tener un 5 en el Examen final para aprobar la asignatura, en caso contrario el alumno deberá presentarse al examen de recuperación. Existe la posibilidad de aprobar el examen con un 4.5 si el alumno ha participado de forma activa en el foro de la asignatura.

Fechas de Evaluación

- ▶ Examen parcial (no hay)
- ▶ Examen final el 21 de Enero de 2021
- ▶ Recuperación 1 de Febrero de 2021

Session info

sessionInfo()

R version 4.0.2 (2020-06-22)

Platform: x86_64-w64-mingw32/x64 (64-bit)

Running under: Windows 10 x64 (build 18362)

Matrix products: default

locale:

[1] LC_COLLATE=Spanish_Spain.1252 LC_CTYPE=Spanish_Spain.1252

[3] LC_MONETARY=Spanish_Spain.1252 LC_NUMERIC=C

[5] LC_TIME=Spanish_Spain.1252

attached base packages:

[1] stats graphics grDevices utils datasets methods base

loaded via a namespace (and not attached):

[1] compiler_4.0.2 magrittr_1.5 tools_4.0.2 htmltools_0.5.0

[5] yaml_2.2.1 stringi_1.4.6 rmarkdown_2.3 knitr_1.29

[9] stringr_1.4.0 xfun_0.16 digest_0.6.25 rlang_0.4.7

[13] evaluate_0.14