**Economía Computacional**

**Primavera 2021**

**TEMARIO GENERAL**

Isidoro Garcia Urquieta

EVALUACION DEL CURSO

Tareas 50%

Midterm 25%

Proyecto Final 25%

**Obligatorio ir con una laptop a la clase.**

OBJETIVOS DE LA ASIGNATURA

En este curso los alumnos aprenderán técnicas estadísticas y econométricas relacionadas con Big Data. Los alumnos podrán explorar y analizar bases de alta dimensionalidad para construir e interpretar modelos predictivos. Más aún, se darán aplicaciones que combinan dichos algoritmos con Teoría Económica en Negocios, Finanzas y Políticas Públicas. Con esto se espera que el alumno pueda aplicar sus conocimientos teóricos a la resolución de diversos problemas reales.

Las técnicas cubiertas incluyen aprendizaje supervisado y no supervisado. Se aprenderá: Regresión, Regularización, Criterios de Información (AIC, BIC), Re-muestreo (Bootstap y Cross Validation), Clasificación (Logística y Árboles de decisión), Clusterización (K-means y PCA), Experimentación, Inferencia Causal, Causal Machine Learning y Natural Language Processing. Se pondrá énfasis en que se aprenda como aplicar estas herramientas en ejemplos reales y cómo decidir qué herramienta es la adecuada.

BIBLIOGRAFIA

* Referencias:
  1. Taddy, Matt “Business Data Science: Combining Machine Learning and Economics to Optimize, Automate and Accelerate Business Decisions”. McGraw-Hill, Ed 2019.
  2. Gareth, James, Daniela Witten, Trevor Hastie and Robert Tibshirani. “An Introduction to Statistical Learning with Applications in R”. Springer 2013.

Software utilizado

* Las clases, tareas y examen se harán en R (<https://www.r-project.org> ). Se recomienda a los alumnos tener descargado e instalado R antes de iniciar el curso.
* Adicionalmente, les recomiendo descargar Rstudio (https://rstudio.com/products/rstudio/download/) . Este es un IDE (Integrated Development Enviroment) para R. Esto es un software que permite interactuar con R de una manera más amigable.
* No se asume un conocimiento previo de R. En las clases y tareas les daré guía sobre estrategias de programación en R. **Sin embargo, esta no es una clase de R.** Les recomiendo familiarizarse con el lenguaje antes de clase para poder enfocarnos en aprender las técnicas de Data Science.
* Algunas buenas referencias para aprender R:
  1. Wickman Hadley, Garret Grolemund, R for Data Science, O’ Reilly. Disponible en línea en <https://r4ds.had.co.nz/index.html> (R4DS)
  2. Google: Intenta buscar ‘How to do XX in R”.
  3. La mayoría de las respuestas a 2 vienen en Stackoverflow ([www.stackoverflow.com](http://www.stackoverflow.com)) o R-bloggers (<https://www.r-bloggers.com>).
  4. Para preguntar cosas más relacionadas a estadística ir a Cross Validated (<https://stats.stackexchange.com>).
* Finalmente, no se dará soporte a tareas y trabajos entregados en Python. Python es un gran lenguaje de programación; pero entenderán la dificultad de enseñar Data Science aunado a dos lenguajes de progamación.

Repositorio del curso: Github

* Las clases, tareas y entregas se harán en el repositorio del curso en Github (<https://github.com/isidorogu/ITAM_Eco_computacional>). Git y Github son herramientas de control de versiones muy utilizadas en el mundo de Data Science y es crucial que las aprendan para ser Científicos de Datos exitosos.
* **¿Qué es Git?** Es un lenguaje de control de versiones de código open source para el desarrollo de software. Es un software elemental para equipos que colaboran con código.
* **¿Qué es Github?** Es n
* .

SESIONES

Habrá 17 sesiones durante el semestre.

|  |  |  |
| --- | --- | --- |
| Sesión | Fecha | Descripción y Objetivos |
| 1 | 12/01/2021 | Inferencia en Big Data: FDR, Bonferroni Test |
| 2 | 19/01/2021 | Regresión Lineal |
| 3 | 26/01/2021 | Selección de Modelos: AIC, BIC, RMSE, LASSO, Ridge |
| 4 | 02/02/2021 | Clasificación: Binaria, Multinomial, Logit |
| 5 | 09/02/2021 | Árboles de decisión: Regresión y Clasificación  Bagging y Boosting |
| 6 | 16/02/2021 | Árboles de decision 2: XgBoosting y Random Forests |
| 7 | 23/02/2021 | Clustering: K-means, Hierarchical Clustering, Topic Models |
| 8 | 02/03/2021 | Factor Models: PCA, PCR, PLS |
| 9 | 09/03/2021 | Text Analysis |
| 10 | 16/03/2021 | Inferencia Causal 1: Experimentación |
| 11 | 23/03/2021 | Inferencia Causal 2: Métodos Cuasiexperimentales |
| 12 | 06/04/2021 | Causal Machine Learning |
| 13 | 13/04/2021 | JAM |
| 14 | 20/04/2021 | Gerry |
| 15 | 27/04/2021 | Armengol |
| 16 | 04/05/2021 | Examen Final |
| 17 | 11/05/2021 | Git: Control de versiones (Base Pricing) |