**Economía Computacional**

**Primavera 2023**

**TEMARIO GENERAL**

Isidoro García Urquieta

EVALUACION DEL CURSO

Tareas 50%

Quizzes 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, no supervisado, inferencia y causalidad. Se aprenderá: Regresión, Regularización, Criterios de Información (AIC, BIC), Re-muestreo (Bootstrap 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.
  3. Imbens, Guido, Donald B.Rubin, “Causal Inference for Statistics, Social, and Biomedical Sciences”. Cambridge University Press. Ed 2017
  4. Hastie, Trevor, Robert Tibshirani and Jerome Friedman, “The Elements of Statistical Learning. Data Mining, Inference and Prediction”. Springer 2017
  5. Chan, Felix and Laszlo Matyas, “Econometrics with Machine Learning”. Springer 2022

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.
  1. En las tareas también subiré las respuestas en Python
* 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.
* En Python, les recomiendo descargar la versión de Anaconda mas un IDE apropiado (Jupyter Notebooks, Vstudio, Rstudio, Pycharm)
* 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. swirl::swirl(). Esta es una librería en R que te permite aprender R en R.
  5. Para preguntar cosas más relacionadas a estadística ir a Cross Validated (<https://stats.stackexchange.com>).

Github

* **¿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 un IDE para git. Te permite interactuar con el control de versiones de una manera amigable. Les recomiendo abrir cuenta (gratuita) en Github. En el mundo de Data Science es muy común que visites los repositorios de otros para buscar inspiración. Eso lo convierte además en un mecanismo reputacional.

.

Temario por clase

Habrá 17 sesiones durante el semestre.

|  |  |  |
| --- | --- | --- |
| Sesión | Fecha | Descripción y Objetivos |
| 1 | 10/01/2023 | Intro & Git |
| 2 | 17/01/2023 | Inferencia en Big Data: FDR, Bonferroni Test |
| 3 | 24/01/2023 | Regresión Lineal |
| 4 | 31/01/2023 | Selección de Modelos: AIC, BIC, RMSE, LASSO, Ridge |
| 5 | 07/02/2023 | Clasificación: Binaria, Multinomial, Logit |
| 6 | 14/02/2023 | Árboles de decisión: Regresión y Clasificación |
| 7 | 21/02/2023 | Árboles de decision 2: Bagging, Boosting, XGBoosting y Random Forests |
| 8 | 28/03/2023 | Clustering: K-means, Hierarchical Clustering, Topic Models |
| 9 | 07/03/2023 | Factor Models: PCA, PCR, PLS, Text Analysis |
| 10 | 14/03/2023 | Survey Design & Analysis |
| 11 | 28/03/2023 | Inferencia Causal 1: Experimentación |
| 12 | 11/04/2023 | Inferencia Causal 2: Métodos Cuasiexperimentales |
| 13 | 18/04/2023 | Causal Machine Learning |
| 14 | 25/04/2023 | Causal Machine Learning II: Efectos heterogeneos |
| 15 | 02/05/2023 | Presentación Proyecto final |
| 16 | 09/05/2023 | Presentación Proyecto Final |