

Teachers: (i) Roberto Rigobon
rigobon@mit.edu
MIT

(ii) Alvaro Riascos
alvaro.riascos@quantil.com.co
Universidad de los Andes

ESCUELA DE VERANO 2019

Salón: Pendiente

Fechas: 25 de junio al 5 de julio. Incluye clase el sábado 29 de junio

Lunes a viernes 8:00 am a 12:00 pm

Sábado 8:00 am a 12:00 pm

Programa sujeto a cambios

OBJETIVOS DE LA MATERIA

Este curso introduce a los estudiantes a algunas de las aplicaciones más importantes de la minería de datos y el aprendizaje de máquinas (Big Data) a problemas de políticas públicas, economía aplicada y macroeconomía. Como veremos, los economistas tienen bastante que aprender de los ingenieros y los ingenieros tienen bastante que aprender de los economistas, y en este curso usaremos una estrategia que compile lo mejor de los dos mundos. Además de aprender los fundamentos teóricos, los estudiantes aprenderán herramientas básicas de programación en R o Python, lo que les permitirá una vez finalizado el curso, plantearse un problema que utilice diversas fuentes de datos y aplicar las técnicas aprendidas en clase para ofrecer una solución novedosa. La primera parte del curso tiene como objetivo aprender los fundamentos teóricos del aprendizaje estadístico como marco teórico de la minería de datos (el problema de aprendizaje máquinas, el compromiso entre sesgo y varianza, aproximación y error, riesgo, consistencia, regularización, complejidad, etc.). Paralelamente se van aprender las principales técnicas de minería de datos (método de vecindades, árboles, boosting, máquinas de vectores de soporte, conglomerados, reglas de asociación, validación cruzada, regularización, etc.) y una introducción a las redes neuronales y el aprendizaje profundo. En la primera semana de clase nos enfocaremos en aplicaciones a políticas públicas: modelos de predicción del crimen, problemas de salud pública, educación y problema en la rama judicial. La segunda semana del curso, exploraremos las soluciones a los problemas de identificación que tanto preocupan a los economistas y se establecerá un puente entre los temas de la primera semana y la teoría económica: predicción, causalidad, identificación y econometría, en donde el énfasis de las aplicaciones será a temas de macroeconomía.

CONTENIDO

Semana	Tema	Expositor	Rerefencia
1	Fundamentación estadística del aprendizaje de máquinas y la minería de datos. Aplicación: Breve introducción al Aprendizaje de máquinas en economía: predicción vs. causalidad, diseño experimental, datos panel y llenado de matrices y estimación de demanda	Alvaro Riascos	[LS] [JWHT]: Capítulo 1,2., [HTF]: Capítulo 1,2 Aplicaciones: Athey, Imbens [2019]. Kleinberg, Ludwig, Mullainathan, Obermeyer [2015]
2	Métodos Lineales de clasificación (regresión lineal, modelo logístico) y tópicos: regularización, validación cruzada, (sub) bagging, boosting, curva ROC. Aplicación: Sesgo, discriminación, riesgo de crédito y riesgo moral	Alvaro Riascos	[JWHT]: Capítulos 3,4,5,6 Aplicaciones: Fuster, Goldsmith-Pinkham, Ramadorai, Walther. [2018]. Sendhil Mullainathan and Ziad Obermeyer. [2017]
3	Métodos no lineales: kernels, naive Bayes árboles, bosques aleatorios, Boosting. Aplicación: Predicción del crimen y sesgos de retroalimentación	Alvaro Riascos	[JWHT]: Capítulo 7,8. Aplicaciones: Dulce, Ramirez y Riascos [2018] Lum, Isaac [2016]. Ensign, Friedler, Scheidegger, Venkatasubramanian. [2018]
4	Análisis no supervisado: Clustering (kmedias), reglas de asociación, minería de texto. Aplicación: Detección automática de tópicos en textos (Gapmaps), Índice de confianza al consumidor en tiempo real	Alvaro Riascos	[JWHT]: Capítulo 10 [HTF]: Capítulo 14 Aplicaciones: Por definir
5	Introducción a aprendizaje profundo (Deep Learning) Aplicación: Riesgo de crédito, estimación de actividad económica usando imágenes satelitales	Alvaro Riascos	Catherine F. Higham, Desmond J. Higham. [2018] Aplicaciones: Por definir
6	Medición, Macroeconomía y las grandes oportunidades de BigData *Data Diseñada y Data Orgánica *Pilares de las mediciones futuras *Debilidades y retos de Big Data	Roberto Rigobon	Notas de clases serán distribuidas después de las clases de esta semana
7	Problemas de Identificación en Macroeconomía: Soluciones Preliminares	Roberto Rigobon	
8	Soluciones Part II. Aplicaciones: Problema de atribución y Problema de Contagio	Roberto Rigobon	
9	Uniendo BigData y Econometría. ¿Qué puede aprender cada uno del otro?	Roberto Rigobon	
10	Aplicaciones: Medición de Estrés, Discriminación, Tráfico Humano, Consumo de Drogas, e Impacto Ambiental	Roberto Rigobon	

METODOLOGÍA

Este curso es muy práctico y requiere de la participación intensa de los estudiantes para su desarrollo. Las clases magistrales tendrán una duración de tres horas complementadas con una hora adicional de programación en Python o R. No es necesario saber de programación en estos lenguajes aunque es muy deseable haber estado expuesto a un nivel introductorio cualquier software estadístico (Stata, SPSS, EViews, etc) o lenguaje de programación (Matlab, Mathematica, etc).

Los estudiantes tendrán que formar grupos (de mínimo dos personas a máximo cuatro personas) para hacer las siguientes entregas:

- Taller 1 (30% de la nota)
- Taller 2 (30% de la nota)
- Prepropuesta de proyecto final (10% de la nota)
- Proyecto final (30% de la nota)

SISTEMA UTILIZADO PARA APROXIMAR LA NOTA DEFINITIVA

El sistema de notas definitivas es el siguiente: las notas totales con decimales en 0 o en .5 no se modificarán. Las notas totales con decimales entre .25 a .49 y entre .75 a .99, se aproximarán a la nota definitiva siguiente. Las notas con decimales entre .01 a .24 y entre .51 a .74, se aproximarán a la nota definitiva anterior.

BIBLIOGRAFÍA

Las referencias principales del curso son:

- [LS]: Luxburg, U., B. Scholkopf. 2008. Statistical Learning Theory: Models, Concepts and Results. <http://arxiv.org/abs/0810.4752>
- [JWHT]: Introduction to Statistical Learning with Applications in R. <http://www-bcf.usc.edu/~gareth/ISL/>
- [HTF]: Hastie, T., Tibshirani, R. y J. Hastie. 2009. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Segunda Edición. Springer http://web.stanford.edu/~hastie/local.ftp/Springer/OLD/ESLII_print4.pdf
- Athey, S., and G. Imbens. 2019. Machine Learning Methods Economists Should Know About.
- Dulce, M., Ramirez, S. y A. Riascos (2018). Efficient allocation of law enforcement resources using predictive police patrolling. <https://arxiv.org/pdf/1811.12880.pdf>
- Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. 2015. Prediction Policy Problems

- Lum, Isaac. 2016. To predict and to serve?
- Ensign, Friedler, Scheidegger, Venkatasubramanian. 2018. Runaway Feedback Loops in Predictive Policing
- Fuster, Goldsmith-Pinkham, Ramadorai, Walther. 2018. Predictably Unequal? The Effects of Machine Learning on Credit Markets.
- Sendhil Mullainathan and Ziad Obermeyer. 2017. Does Machine Learning Automate Moral Hazard and Error?
- Catherine F. Higham, Desmond J. Higham. 2018. Deep Learning: An Introduction for Applied Mathematicians.

INFORMACIÓN ADICIONAL

El estudiante podrá retirar el curso, sin devolución, hasta un día hábil antes de la fecha del examen final estipulado por el Profesor. La Universidad no devolverá el dinero cancelado por concepto de matrículas de estos cursos de la Escuela de Verano.