

Profesor: Alvaro J. Riascos Villegas
 Contacto: e-mail: ariascos@uniandes.edu.co
 Oficina: Bloque W, Oficina 918.
 Página en Internet del curso: www.alvaroriascos.com

1. Objetivos de la materia

Este seminario introduce los estudiantes en los pilares teóricos fundamentales de la teoría de aprendizaje estadístico como marco teórico de la minería de datos (el problema de aprendizaje o *machine learning*, el compromiso entre sesgo y varianza, aproximación y error, riesgo, consistencia, regularización, complejidad, etc.). Paralelamente se van a aprender las principales técnicas de minería de datos (método de vecindades, redes neuronales, redes bayesianas, árboles, *boosting*, *cross validation*, máquinas de vectores de soporte, *clustering*, reglas de asociación, minería de texto, etc.) a través de ejemplos y problemas que los estudiantes deberán implementar y resolver haciendo uso de un computador y, en lo posible, resolviendo problemas que sean de su interés (análisis de datos de redes sociales, reconocimiento de caracteres, extracción de señales, pronósticos, clasificación de clientes, alertas de fraude, anomalías, etc.).

Este curso es muy práctico y requiere de la participación activa de los estudiantes, haciendo talleres de programación y un trabajo final de investigación. Para facilitar la aplicabilidad de las técnicas aprendidas habrá una introducción corta a R (<http://www.r-project.org/>) y parte de las técnicas se aprenderán de forma paralela con su implementación en este lenguaje de programación y sus respectivos paquetes.

2. Contenido

Semana	Tema	Referencias
4	Aprendizaje Estadístico I: Modelos, Conceptos y Resultados	[JWHT]: Capítulo 1. [HTF]: Capítulo 1.
5	Aprendizaje Estadístico II: Modelos, Conceptos y Resultados	[LS] [JWHT]: Secciones 2.1, 2.2.
6	Aprendizaje Estadístico III: Modelos, Conceptos y Resultados Entrega Taller I	[LS] [JWHT]: Secciones 2.1, 2.2.
7	Métodos lineales de clasificación Aplicaciones	[JWHT]: Capítulo 3,4. [HTF]: Capítulo 3,4.
8	Validación y Selección de Modelos Entrega Taller II	[HTF]: Capítulos 7 y 8

		[JWHT]: Capítulo 5
9	Métodos no lineales: Árboles, Boosting de Árboles y Random Forests	[HTF]: Capítulos 9, 10 y 15
10	Análisis no supervisado I Entrega Taller III	[HTF]: Capítulos 14
11	Análisis no supervisado II	[HTF]: Capítulos 14
12	Aprendizaje de máquinas y discriminación: Dilemas éticos	
13	Receso Última fecha 30%	
14	Aprendizaje de máquinas y causalidad	
15	Semana Santa	
16	Redes neuronales profundas	
17	Redes Convolucionales y Recurrentes Entrega Taller IV	
18	No hay clase	
19	Encoders y GANs Propuesta trabajo final	
20	Minería de Texto I	
21	Minería de Texto II	

3. Metodología

Este curso es muy práctico y requiere de la participación intensa de los estudiantes para su desarrollo. Los estudiantes tendrán que formar grupos (de mínimo dos personas a máximo cuatro personas) para hacer las siguientes entregas:

- Taller 1 (15% de la nota)
- Taller 2 (15% de la nota)
- Taller 3 (15% de la nota)
- Taller 4 (15% de la nota)
- Propuesta de proyecto final (10% de la nota).
- Proyecto final (30% de la nota): Máximo seis páginas incluyendo tablas, gráficos, bibliografía, etc.

4. 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.

5. Referencias

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

Las referencias de las aplicaciones se irán subiendo a la página web del curso a lo largo del semestre.

6. Lecturas¹

Part I: Equality of Opportunity

Geography of Economic Mobility

Chetty, Raj, John Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter. 2018. "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility." NBER Working Paper No. 25147. [Non-technical summary.](#)

Historical Trends

Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. 2017. "The Fading American Dream: Trends in Absolute Income Mobility Since 1940." *Science* 356 (6336): 398-406. [Non-technical summary](#)

¹ Tomado de Using Big Data to Solve Economic and Social Problem. Profesores Raj Chetty y Greg Bruich de la Universidad de Harvard.

Part II: Education

Higher Education

Dynarski, Susan, C.J. Libassi, Katherine Michelmore, and Stephanie Owen. 2018. "Closing the Gap: The Effect of a Targeted, Tuition-Free Promise on College Choices of High-Achieving, Low-Income Students." NBER Working Paper No. 25349

Primary Education

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2011. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104 (9): 2633–79. [Non-technical summary.](#)

Part IV: Health

Improving Health Outcomes

Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. "The Parable of Google Flu: Traps in Big Data Analysis." *Science* 343 (6167): 1203–5.

The Economics of Health Care and Insurance

Finkelstein, Amy, Nathaniel Hendren, and Mark Shepard. 2017. "Subsidizing Health Insurance for Low-Income Adults: Evidence from Massachusetts." NBER Working Paper No. 23668. [Non-technical summary](#)

Part V: Criminal Justice

Heller, Sara B., Anuj K. Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, Harold A. Pollack. 2015. "Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago." NBER Working Paper No. 21178.

Hvistendahl, Mara. 2016. [Can 'Predictive Policing' Prevent Crime Before It Happens?](#) *Science News*.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017. "Human Decisions and Machine Predictions." NBER Working Paper No. 23180.

Kleinberg, John, Jens Ludwig, and Sendhil Mullainathan. 2016. [A Guide to Solving Social Problems with Machine Learning](#). *Harvard Business Review*.

Part VI: Climate Change

Policies to Mitigate Climate Change

Ito, Koichiro. 2014. "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing." *American Economic Review* 104 (2): 537–63.

Li, Shanjun, Joshua Linn, and Erich Muehlegger. 2014. "Gasoline Taxes and Consumer Behavior." *American Economic Journal: Economic Policy* 6 (4): 302–42

Schultz, P. Wesley, Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vidas Griskevicius. 2007. "The Constructive, Destructive, and Reconstructive Power of Social Norms." *Psychological Science* 18 (5): 429–34.

Part VIII: Economic Development and Institutional Change

Dell, Melissa. 2010. "The Persistent Effects of Peru's Mining Mita." *Econometrica* 78(6): 1863-1903.