

1 Profesores y horarios

Profesor: Ignacio Sarmiento-Barbieri (i.sarmiento@uniandes.edu.co)

- Horario Clase: Martes y Jueves, 2:00 p.m. – 3:15 p.m. (Aula virtual Zoom)
- Web del Curso: <https://github.com/ECON-4676-UNIANDES-Fall-2021>
- Horario de atención a estudiantes: Via Slack

Profesor Complementario: Rafael Cano Polania (r.cano@uniandes.edu.co)

- Horario Clase: Viernes, 15:30 p.m. – 16:45 p.m. (Aula virtual Zoom)
- Horario de atención a estudiantes: Via Slack, Miércoles 6:30 a.m- 7:30 a.m

Cláusula de ajustes razonables

Si lo considera pertinente, siéntase en libertad de informar al profesor lo antes posible **si usted tiene alguna condición, visible o invisible**, por la cual requiera algún ajuste para estar en igualdad de condiciones con los y las demás estudiantes. Debido a las actuales circunstancias, las barreras de conectividad o acceso a los recursos tecnológicos indispensables para la clase son parte de las condiciones que pueden requerir ajustes. Por la misma razón, no necesitará presentar documentación para solicitar esos ajustes.

También lo invitamos a buscar asesoría y apoyo en la Coordinación de su programa, en la Decanatura de Estudiantes (<http://centrodeconsejeria.uniandes.edu.co>, Bloque Ñf, ext. 2207, 2230 y 4967, horario de atención L-V 8:00 a.m. a 5:00 p.m.) o en el Programa de Acción por la Igualdad y la Inclusión Social (PAIIS) de la Facultad de Derecho (paiis@uniandes.edu.co). Si su solicitud se basa en dificultades de acceso a conectividad o tecnología, es particularmente importante que haga este contacto adicional para que pueda acceder a los recursos de apoyo que brinda la Universidad.

Se entiende por ajustes razonables todas “las modificaciones y adaptaciones necesarias y adecuadas que no impongan una carga desproporcionada o indebida, cuando se requieran en un caso particular, para garantizar a las personas con discapacidad el goce o ejercicio, en igualdad de condiciones con las demás, de todos los derechos humanos y libertades fundamentales” Convención sobre los Derechos de las personas con discapacidad, art.2.

Si quiere más información sobre ajustes razonables, puede visitar <https://agora.uniandes.edu.co/que-son-los-ajustes-razonables/>. Y sobre la política de momentos difíciles, <https://agora.uniandes.edu.co/sabes-que-es-la-politica-de-momentos-dificiles/>.

Cláusula de respeto por la diversidad y contra el acoso

Todos debemos respetar los derechos de quienes hacemos parte de esta comunidad académica. En esta comunidad consideramos inaceptable cualquier situación de acoso, acoso sexual, discriminación, matoneo, y/o amenaza. La persona que se sienta en alguna de estas situaciones puede denunciar su ocurrencia y buscar orientación y apoyo ante alguna de las siguientes instancias:

- El equipo pedagógico de este curso, la Coordinación o la Dirección del programa de Economía.
- La Decanatura de Estudiantes (DECA, Ed. Ñf-Casita amarilla).
- La Ombudsperson (ombudsperson@uniandes.edu.co, Edificio RGA–Pedro Navas, Of. 201, ext. 5300 y 3933).
- El Comité MAAD (lineamaad@uniandes.edu.co, <https://uniandes.edu.co/MAAD> o a la ext. 2707 o 2230). Si quieren mayor información, guía o necesitan activar el protocolo MAAD pueden acudir a Nancy García (n.garcia@uniandes.edu.co) en la Facultad. Para mayor información sobre el protocolo MAAD, puede visitar esta página: <https://decanatura.deestudiantes.uniandes.edu.co/index.php/es/sobre-la-decanatura/827>
- Grupos de apoyo estudiantiles que pueden ofrecerle apoyo y acompañamiento: No Es Normal (derechoygenero@uniandes.edu.co o <https://www.facebook.com/noesnormaluniandes/?fref=ts>); Pares de Acompañamiento Contra el Acoso-PACA (paca@uniandes.edu.co o <https://www.facebook.com/PACA-1475960596003814/?fref=ts>).
- Para mayor información sobre el protocolo MAAD, puede visitar esta página: <https://agora.uniandes.edu.co/wp-content/uploads/2020/09/ruta-maad.pdf>

2 Introducción y descripción general del curso

El curso busca que los estudiantes sean capaces de organizar, relacionar, y analizar datos anárquicos y espontáneos de gran escala, tipo, y calidad para responder preguntas económicas. Problemas de predicción e inferencia, con especial énfasis en inferencia causal, atravesarán transversalmente al curso. Mediante una combinación de talleres, presentaciones, y un trabajo final grupal, los estudiantes adquirirán las herramientas estadísticas y computacionales necesarias para hacer uso de big data y machine learning en investigación empírica.

Es pre-requisito haber cursado Microeconomía 3, Econometría 1 y 2. Se recomienda haber cursado Econometría Avanzada o equivalente. Se aconseja tener experiencia con programación en R o Python, pero no es requisito. ¡Estos programas (y todos) se aprenden utilizándolos!

3 Objetivos y competencias del curso

El objetivo de este curso es introducir a los alumnos a un conjunto de herramientas estadísticas, matemáticas, y computacionales para abordar problemas de gran cantidad/tipos/calidad de datos (“large n”), y cantidad de variables (“large p”). Problemas de predicción e inferencia, con especial énfasis en inferencia causal, atravesarán transversalmente al curso. Se buscará también familiarizar a los alumnos con la literatura reciente que utiliza estas herramientas.

Competencias

- Comprender las técnicas provenientes de la ciencia de datos, la ciencia computacional, y la estadística desde una visión de economistas
- Ser capaz de contrastar distintas técnicas econométricas y su conveniencia para contestar preguntas económicas y sociales.
- Desarrollar habilidades técnicas para el manejo cuantitativo de datos que surgen de distintas fuentes: texto, geoespaciales, encuestas, paginas web, etc.
- Desarrollar la capacidad manejar, analizar y sintetizar bases de datos con gran numero de observaciones y variables para generar conclusiones y recomendaciones sobre preguntas relevantes a las ciencias sociales.
- Aprender a manejar con fluidez distintas herramientas computacionales.
- Capacidad de analizar críticamente los datos presentados, ya sean provenientes de textos académicos o de los ejercicios presentados en el cursado
- Trabajar y resolver situaciones en grupo
- Habilidad de exponer y defender (en forma escrita y oral) el trabajo realizado

4 Organización del curso

Table 1: Temario y Organización (sujeto a cambios)

Semana	Temas de clase	Actividades
1	Introducción al curso. Propiedades Numericas de OLS. FWL	
2	Computación de OLS; Maximum Likelihood; Optimización. Git y SQL. Programación en paralelo	Problem Set 1
3	Estadística Bayesiana.	
4	Estadística Bayesiana.	Problem set 2
5	Datos y Modelos Espaciales	
6	Modelos espaciales. Introducción a Machine Learning.	
7	Sobreajuste. Validación cruzada. Selección de modelos.	Problem set 3
8	Regularización (Lasso, Ridge).	
9	Regularización in Econ (Belonni). Modelos de clasificación.	Problem set 4
	Receso	
10	Modelos de clasificación. Imbalanced learning.	Entrega de propuesta escrita trabajo final
11	Árboles de decisión (CARTs, Random Forest, Bagging y aplicaciones para inferencia causal)	
12	Boosting: XGboost, LightGBM	Problem set 5
13	Aprendizaje no Supervisado. Modelos de clusterización y aplicaciones a texto. Texto como Datos	
14	Reducción de dimensión: PCA, Modelos de categorización de tópicos.	Problem set 6
15	Word Embeddings. Introducción a aprendizaje profundo. Redes neuronales. Imagenes como datos	
16	Presentaciones finales	
17	Semana de finales	Entrega proyecto final

5 Metodología

La metodología del curso combina clases virtuales, talleres, y un proyecto final. La participación de los estudiantes es fundamental para sacar el mayor provecho del curso. La virtualidad impone nuevos desafíos y es importante mantenerse conectados para crear las sinergias que surgen de las interacciones humanas. A su vez, los estudiantes realizarán trabajos prácticos grupales para evaluar su aprendizaje.

El producto final de este curso es proyecto implementando los conceptos y herramientas aprendidas a un problema concreto. El proyecto puede estar orientado a la investigación o a la industria dependiendo los intereses del alumno

6 Criterios de evaluación

Table 2: Puntajes

	Puntaje Individual	Puntaje Total
Participación		10%
Talleres	10%	60%
Trabajo Final		30%
Primer Entrega	5%	
Presentaciones	10%	
Entrega Final	15%	
Total		100%

- Participación. Si bien participación es la actividad con menos peso en la composición final, será el “tiebreaker” por el cual decidire la nota final. Participación no incluye solamente la asistencia a clases, sino también actividades fuera de clase. Una vez registrados en el curso los estudiantes recibirán invitación al canal de Slack, a GitHub, y a la nube. La participación será juzgada en función a la participación en las discusiones, en los trabajos grupales, de las interacciones en el canal de Slack, el aprovechamiento de la nube. Se espera que estudiantes encuentren al menos tres errores de tipeo o cualquier otro tipo y los arreglen a través de pull requests en GitHub.
- Talleres. Los estudiantes realizarán trabajos prácticos grupales para evaluar su aprendizaje. Los grupos no podrán superar los 3 miembros. Habrá 6 talleres durante el semestre. Se dedicarán al menos 6 clases para la discusión y presentación de los talleres. Los talleres serán entregados vía GitHub. Parte de nota de la participación saldrá de la evaluación de la historia del repositorio donde se verá la contribución de cada estudiante.
- Trabajo final. La actividad estará dividida en 3 entregas. En la primera entrega los grupos entregaran una propuesta escrita donde se presenta la idea y como planean llevarla a cabo.

En una segunda entrega donde se expondrán los avances y resultados. La entrega final será al concluir el curso que consolida todo el trabajo.

7 Notas definitivas, reclamos y fraude académico

Sistema de aproximación de notas definitiva

De acuerdo con el Artículo 51 del Reglamento General de Estudiantes de Pregrado, “Las calificaciones definitivas de las materias serán numéricas de uno cinco (1,5) a cinco (5,0), en unidades, décimas y centésimas.” En este curso se aproximará la nota a la centésima más cercana. Por ejemplo, si el cálculo del cómputo es 3.245, la nota final se aproximará a 3.25; si el resultado del cálculo es 2.994 la nota final será de 2.99

Excusas

Los estudiantes que no presenten las actividades y evaluaciones del curso en la fecha establecida previamente recibirán una calificación de cero (0), a menos que justifiquen su ausencia ante el profesor dentro de un término no superior a ocho (8) días hábiles. Se consideran excusas válidas aquellas contempladas en el artículo 43 del Reglamento General de Estudiantes de Pregrado. Si la excusa presentada por el estudiante es válida, el profesor programará una evaluación o actividad supletoria en el transcurso de las dos semanas siguientes.

Ausencias en días de trabajo en grupo (clase tipo X)

La asistencia a clase en los días donde hay trabajo en grupo es obligatoria. Si el estudiante se ausenta deberá presentar una excusa válida. Si no presenta una excusa válida su nota en la evaluación Tipo X se verá reducida. En particular su nota θ (ver página 5) de la actividad correspondiente al día de su ausencia se reducirá en 20 puntos por cada día de ausencia. Ver la sección anterior, “Excusas”, para la definición de excusa válida.

Reclamos

De acuerdo con los artículos 64 y 65 del Reglamento General de Estudiantes de Pregrado: “Todo estudiante que desee formular un reclamo sobre las calificaciones de cualquier evaluación o sobre la nota definitiva del curso, deberá dirigirlo por escrito y debidamente sustentado al profesor responsable de la materia, dentro de los cuatro (4) días hábiles siguientes a aquel en que se dan a conocer las calificaciones en cuestión. El profesor dispone de cinco (5) días hábiles para resolver el reclamo formulado; vencido el término informará al estudiante la decisión correspondiente.” “Si el estudiante considera que la decisión no corresponde a los criterios de evaluación, podrá solicitar la designación de un segundo calificador mediante un escrito debidamente sustentado, dirigido al Consejo de Facultad o de Departamento, según el caso, dentro de los cuatro (4) días hábiles siguientes al conocimiento de la decisión. Si el Consejo encuentra fundada la solicitud, procederá a

designar, solamente para tal efecto, un segundo calificador cuya decisión debidamente sustentada será definitiva e inmodificable. En ningún caso, el segundo calificador podrá desmejorar la nota inicialmente asignada por el profesor.”

Fraude

Este curso se acoge a la reglamentación sobre faltas académicas y disciplinarias consignada en el capítulo 10 del Reglamento General de Estudiantes de Pregrado de la Universidad. El fraude se considera como una de las faltas más graves que se puedan cometer en la Academia. En una comunidad académica la discusión respetuosa e informada de las ideas propias y ajenas constituye la columna vertebral de su quehacer diario. No dar crédito a otros por sus ideas publicadas o no, o apropiarse de las ideas de otros, sean estos autores, profesores o compañeros, atenta directamente contra la esencia de la Academia. En consecuencia, en este curso, el fraude en cualquiera de sus formas, es considerado como una acción inaceptable.

8 Fechas Importantes

- Inicio de clases: 9 de agosto.
- Semana de receso: 4-9 de octubre.
- Fecha de entrega del 30% de las notas en MiBanner: 15 de Octubre.
- Último día de clases: 4 de diciembre.
- Exámenes finales (solo para cursos con varias secciones y exámenes conjuntos): 6-11 de diciembre.
- Último día para solicitar retiros: 17 de diciembre 6 p.m.
- Último día para subir notas finales en MiBanner: 16 de diciembre.

9 Virtualidad

Debido a la coyuntura actual de la pandemia las clases del curso serán virtuales. En el curso tendremos las siguientes pautas respecto a la virtualidad. Estas pautas buscan promover el respeto, la construcción de comunidad y la conexión humana entre estudiantes y profesores.

- Las clases se llevarán a cabo en Zoom.
- Las clases virtuales serán sincrónicas.
- Los estudiantes deben tener el micrófono en silencio. Para participar en clase por favor levantar la mano antes.

- Los estudiantes pueden usar libremente el chat.
- Prender la cámara durante clase es obligatorio. Si en alguna sesión el estudiante no puede usar la cámara, deberá escribir un correo a sus profesores explicando por qué no pudo prender la cámara y qué está haciendo para solucionar la situación.
- Grabaremos las clases donde el profesor magistral presente material que después será evaluado. La grabación de la sesión se publicará en la página del curso.

10 Política de Momentos difíciles

Todas las personas pueden pasar por un momento difícil que de alguna manera pueda afectar nuestra vida en la Universidad. Pueden ser problemas en casa, con la pareja, incluso estrés por esta u otra materia. Si usted siente que está pasando por un momento complicado, **sin importar el motivo**, siéntase con la tranquilidad de hablar con la profesora para pedir tiempo o apoyo. Ningún trabajo o entrega puede sobrepasar su salud mental y física. Su bienestar es lo más importante.

11 Bibliografía (sujeta a cambios)

- Albouy, D., Christensen, P., & Sarmiento-Barbieri, I. (2020). Unlocking amenities: Estimating public good complementarity. *Journal of Public Economics*, 182, 104110.
- Anselin, L. (1982). A note on small sample properties of estimators in a first-order spatial autoregressive model. *Environment and Planning A*, 14(8), 1023-1030.
- Anselin, Luc, & Anil K Bera. 1998. "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics." *Statistics Textbooks and Monographs* 155. MARCEL DEKKER AG: 237-90.
- Arbia, G. (2014). *A primer for spatial econometrics with applications in R*. Palgrave Macmillan.
- Ash, E., Chen, D. L., & Ornaghi, A. (2020). Stereotypes in High-Stakes Decisions: Evidence from US Circuit Courts (No. 1256). University of Warwick, Department of Economics.
- Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola (2020) Dive into Deep Learning. Release 0.15.1.
- Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27), 7353-7360.
- Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11, 685-725.

- Belloni, A., Chernozhukov, V., & Hansen, C. (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, 28(2), 29-50.
- Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2), 608-650.
- Bivand, R. S., & Pebesma, E. J. (2020). *Spatial Data Science* (Chapter 8)
- Bivand, R. S., Gómez-Rubio, V., & Pebesma, E. J. (2008). Applied spatial data analysis with R (Vol. 747248717, pp. 237-268). New York: Springer.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264), 1073-1076.
- Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *Advances in neural information processing systems* (pp. 4349-4357).
- Breiman, L. (2001). "Random Forests". In: *Machine Learning*. ISSN: 1098-6596. DOI: 10.1017/CBO9781107415324.004. eprint: arXiv:1011.1669v3.
- Carneiro, A., Guimarães, P., & Portugal, P. (2012). Real Wages and the Business Cycle: Accounting for Worker, Firm, and Job Title Heterogeneity. *American Economic Journal: Macroeconomics*, 4 (2): 133-52.
- Casella, G. (1985). An introduction to empirical Bayes data analysis. *The American Statistician*, 39(2), 83-87.
- Casella, G., & Berger, R. L. (2002). *Statistical inference* (Vol. 2, pp. 337-472). Pacific Grove, CA: Duxbury.
- Charpentier, Arthur (2018). *Classification from scratch, boosting*.
- Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- Chen, T., He, T., & Benesty, M. (2018). *XGBoost Documentation*.
- Chernozhukov, V., Hansen, C., & Spindler, M (2016). hdm: High-Dimensional Metrics R Journal, 8(2), 185-199.

- Chetty, R., & Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility II: County-level estimates. *The Quarterly Journal of Economics*, 133(3), 1163-1228.
- Christensen, P., Sarmiento-Barbieri, I., Timmins C. (2020). Housing Discrimination and the Pollution Exposure Gap in the United States. NBER WP No. 26805
- Clark, M (2018). An Introduction to Text Processing and Analysis with R.Rstudio (2020). Tutorial TensorFlow
- Constantine, P. G., & Gleich, D. F. (2011, June). Tall and skinny QR factorizations in MapReduce architectures. In *Proceedings of the second international workshop on MapReduce and its applications* (pp. 43-50).
- Davidson, R., & MacKinnon, J. G. (2004). *Econometric theory and methods* (Vol. 5). New York: Oxford University Press.
- Dean, J., & Ghemawat, S. (2004). MapReduce: Simplified data processing on large clusters.
- Efron, B., & Hastie, T. (2016). *Computer age statistical inference* (Vol. 5). Cambridge University Press.
- Einav, Liran, and Jonathan D. Levin. The data revolution and economic analysis. No. w19035. National Bureau of Economic Research, 2013.
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1, No. 10). New York: Springer series in statistics.
- Gentzkow, M., & Shapiro, J. M. (2010). What drives media slant? Evidence from US daily newspapers. *Econometrica*, 78(1), 35-71.
- Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1, No. 2). Cambridge: MIT press.
- Green, D. P., & Kern, H. L. (2012). Modeling heterogeneous treatment effects in survey experiments with Bayesian additive regression trees. *Public opinion quarterly*, 76(3), 491-511.
- Greene, W. H. (2003). *Econometric analysis fifth edition*. New Jersey: Prentice Hall.
- Gu, J., & Koenker, R. (2017). Empirical Bayesball remixed: Empirical Bayes methods for longitudinal data. *Journal of Applied Econometrics*, 32(3), 575-599.
- Hayashi, F. (2000). *Econometrics*. 2000. Princeton University Press. Section, 1, 60-69.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.

- Kasy M. (2019). Trees, forests, and causal trees. Mimeo.
- Koenker, R. (2013) Economics 508: Lecture 4. Model Selection and Fishing for Significance. Mimeo
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Ijcai (Vol. 14, No. 2, pp. 1137-1145).
- Kuhn, M. (2012). The caret package. R Foundation for Statistical Computing, Vienna, Austria.
- Lee, K., & Braithwaite, J. (2020). High-Resolution Poverty Maps in Sub-Saharan Africa. arXiv preprint arXiv:2009.00544.
- Leo Breiman. Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statistical Science, 16(3):199–231, 2001b.
- Lovelace, R., Nowosad, J., & Muenchow, J. (2019). Geocomputation with R. CRC Press. (Chapters 2 & 6)
- Lundberg, I (2017). Causal forests. A tutorial in high dimensional causal inference. Mimeo
- McMillen, D., Sarmiento-Barbieri, I., & Singh, R. (2019). Do more eyes on the street reduce Crime? Evidence from Chicagos safe passage program. Journal of urban economics, 110, 1-25.
- Mullainathan, S. and Spiess, J., 2017. Machine learning: an applied econometric approach. Journal of Economic Perspectives, 31(2), pp.87-106.
- Murphy, K. P. (2012). Machine learning: a probabilistic perspective. MIT press.
- Robinson, D. (2017). Introduction to Empirical Bayes: Examples from Baseball Statistics. 2017.
- Rstudio (2020). Tutorial TensorFlow
- Ruud, P. A. (2000). An introduction to classical econometric theory. OUP Catalogue
- Sarmiento-Barbieri, I. (2016). An Introduction to Spatial Econometrics in R.
- Sosa Escudero, W. (2019). Big Data. Siglo Veintiuno Editores
- Taddy, M. (2019). Business data science: Combining machine learning and economics to optimize, automate, and accelerate business decisions. McGraw Hill Professional.
- Tobler, WR. 1979. "Cellular Geography." In Philosophy in Geography, 379–86. Springer.

- Van Loan, C. F., Golub, G. H. (2012). Matrix Computations. United States: Johns Hopkins University Press.
- Varian, Hal R. Big Data: New Tricks for Econometrics. Journal of Economic Perspectives 28, no. 2 (2014): 3-28.
- Voigt, R., Camp, N. P., Prabhakaran, V., Hamilton, W. L., Hetey, R. C., Griffiths, C. M., ... & Eberhardt, J. L. (2017). Language from police body camera footage shows racial disparities in officer respect. Proceedings of the National Academy of Sciences, 114(25), 6521-6526.
- Wasser, L. GIS With R: Projected vs Geographic Coordinate Reference Systems Last Access September 10, 2020
- Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society, Series B.67: pp. 301–320