



Big Data, ML and Econometrics

Horario: Martes de 19.15 a 22hs.

Aula: A101 ([918 2403 0049](tel:91824030049))

Profesor Gabriel Martos Venturini

email: gmartos@utdt.edu

Universidad Torcuato Di Tella

Departamento de Matemática y Estadística

El programa de esta asignatura puede sufrir alguna modificación en su estructura a lo largo del curso a criterio del profesor.

Descripción breve del curso: Discutimos las técnicas más relevantes para resolver problemas de aprendizaje supervisado en Machine-Learning (ML) en el contexto particular del Bigdata. El enfoque del curso es metodológico-práctico: Después de presentar los aspectos fundamentales de cada modelo, resolvemos casos de estudio en Economía, Finanzas y Negocios utilizando el lenguaje de programación R. A continuación algunos ejemplos de casos resueltos en clase:

- Predicción de fuga de clientes de una compañía telefónica con Random Forest.
- Pricing de pólizas de seguros con máquinas de vector soporte.
- Clasificador de imágenes con redes neuronales.
- Estimación y predicción de la inflación utilizando SSA (Singular Spectrum Analysis).

Objetivos del curso:

1. Introducir herramientas de ML que puedan resultar de utilidad en contextos de modelado y predicción de variables económicas y financieras.
2. Aprender a seleccionar y entrenar los modelos de manera apropiada. Interpretar los resultados de la modelización y comprender el alcance y las limitaciones de los modelos.
3. Implementar los modelos vistos en clase en el entorno de programación R.

Breve descripción del temario para el curso:

- 1- **Introducción:** Bigdata y problemas empíricos modernos con datos. Machine Learning y Econometría. Aprendizaje supervisado vs no supervisado. Bias-Variance trade off. Aprendizaje automático y minimización del riesgo empírico. Introducción a R y algunos conceptos generales de programación. Casos prácticos y ejercicios.
- 2- **Aprendizaje supervisado:** *Clasificación y Regresión.* K-vecinos como un primera aproximación a los problemas de aprendizaje supervisado. Técnicas de validación cruzada. Modelos aditivos y técnicas de selección automática de modelos/variables. Modelos aditivos y regularización (Ridge, Lasso y Elastic Nets). Modelos de partición recursiva: CART. Ensamble de modelos: Bagging, Random Forest y Boosting Machines. Métodos de kernel y Máquinas de Vector Soporte. Introducción a la modelización con Redes Neuronales. Modelos gráficos. Aplicaciones y casos de estudio en Economía, Finanzas y Negocios.

- 3– **Miscelánea:** En caso de disponer de tiempo discutiremos también ciertos *tópicos de aprendizaje supervisado en el contexto de series temporales*. Entre ellos los problemas de clasificación con series temporales, los Procesos Gaussianos para modelar datos de naturaleza temporal. Singular Spectrum Analysis para modelar y predecir series. Validación cruzada en el contexto de series. Aplicaciones y casos de estudio en Economía, Finanzas y Negocios.

Criterio de evaluación: Para aprobar la asignatura la/el alumna/o debe presentar un trabajo práctico en donde se aborde algún problema de aprendizaje supervisado utilizando las técnicas vistas en clase. A continuación se dan algunas pautas generales para presentar el TP.

- Fecha límite de presentación: Febrero de 2025.
 - Quienes **NO presenten el trabajo** antes de la fecha límite se considerarán **ausentes** (la nota correspondiente será una U); mientras que quienes presentando el TP no cumplan con un estándar mínimo de calidad* en el la producción del mismo estarán reprobados (nota D).
 - Aspectos a evaluar:
 - Motivación y contextualización del problema.
 - Originalidad en la presentación.
 - Revisión bibliográfica sobre otros casos de estudio análogos.
 - Benchmark comparativo con modelos econométricos.
 - Reproducibilidad de los resultados.
 - Integración de los conocimientos adquiridos.
- * Ver TP de referencia en el campus.

Bibliografía para el curso

Recomendada:

- Notas y material de clases (código R y artículos sugeridos durante la cursada).
- a) Gareth, Hastie, and Tibshirani. *An introduction to statistical learning*. Springer (online).

Complementaria:

- b) Friedman, Hastie, and Tibshirani. *The elements of statistical learning*. Springer (online).
- c) Efron, Tibshirani and Wainwright. *Statistical learning with sparsity*. CRC Press.
- d) Efron and Hastie. *Computer age statistical inference: Algorithms, evidence, and data science*. Cambridge University Press.
- e) Murphy K.. *Machine Learning. A probabilistic perspective*. MIT press.
- f) Rasmussen and Williams. *Gaussian processes for machine learning*. MIT press.
- g) Alpaydin E. *Introduction to machine learning*. MIT press.