



1. DATOS BÁSICOS DEL TFG:

Título: Estudio y aplicación de Inferencia Causal mediante Double Machine Learning con base en el Teorema de Frish-Waugh-Lovell

Descripción general (resumen y metodología):

La inferencia causal se ocupa de entender las relaciones de causa y efecto en distintos escenarios y estudios experimentales. A diferencia del análisis estadístico tradicional, que se centra en identificar patrones de asociación entre variables, la inferencia causal busca responder a preguntas contrafactuales: ¿cómo cambiaría un resultado si alteráramos deliberadamente un factor del sistema? Esta distinción es fundamental, ya que dos variables pueden estar correlacionadas sin que exista una relación causal entre ellas, como puede ocurrir cuando intervienen factores no observados.

La capacidad de inferir relaciones causales sólidas es esencial en contextos donde se requiere tomar decisiones informadas o intervenir en sistemas complejos. Por ejemplo, en medicina, para evaluar la efectividad real de un tratamiento; en políticas públicas, para anticipar el impacto de una reforma; o en economía y negocios, para optimizar estrategias ante variaciones del entorno. Sin un análisis causal riguroso, se corre el riesgo de tomar decisiones basadas en correlaciones espurias, lo que puede conducir a acciones ineficaces o incluso perjudiciales.

El campo de la Inferencia Causal ha evolucionado significativamente, pasando de enfoques filosóficos y estadísticos tradicionales a marcos teóricos formales y operativos. Modelos como el de los resultados potenciales (potential outcomes) y el uso de grafos dirigidos acíclicos (DAGs) han permitido establecer criterios claros para identificar relaciones causales a partir de datos observacionales. No obstante, cuando se trabaja con estructuras complejas o con un elevado número de variables de confusión, estos métodos presentan limitaciones. Para afrontar estos desafíos, se han desarrollado enfoques modernos que incorporan herramientas de aprendizaje automático. Entre ellos destaca el método conocido como Double Machine Learning (DML) [4], que permite estimar efectos causales de forma robusta en entornos de alta dimensionalidad y con posibles relaciones no lineales.

Este método se basa en descomponer el problema de estimación del parámetro causal en dos etapas. En la primera, se ajustan modelos predictivos para las variables que actúan como confusoras, utilizando técnicas flexibles como los métodos de machine learning. En la segunda, se corrige la estimación causal mediante un proceso de ortogonalización, lo que permite eliminar el sesgo introducido por dichas variables. Esta descomposición se inspira en el Teorema de Frisch-Waugh-Lovell [5, 6], que en regresión lineal permite separar los efectos de distintas variables explicativas mediante una proyección ortogonal.

El desarrollo del presente Trabajo Fin de Grado se articulará en varias etapas. En primer lugar, el estudiante realizará una revisión teórica de los fundamentos de la Inferencia Causal, prestando atención a las limitaciones de los enfoques tradicionales frente a problemas con muchas variables de confusión.

A continuación, se abordará el estudio de la técnica Double Machine Learning, tanto desde el punto de vista teórico como desde su implementación práctica. Se analizará en detalle el papel del teorema de Frisch-Waugh-Lovell en la descomposición de los efectos y su relación con la ortogonalización que garantiza la validez de la estimación causal en DML.

En la parte práctica del trabajo, se diseñarán simulaciones controladas utilizando el paquete simcausal en R u otro similar, con el objetivo de generar datos sintéticos a partir de estructuras causales definidas mediante DAGs. Estas simulaciones permitirán validar la técnica estudiada en

entornos donde el efecto causal verdadero es conocido, facilitando así la comprensión del procedimiento y su robustez.

Finalmente, se aplicará la metodología a un conjunto de datos reales. El estudiante deberá realizar un análisis completo del problema, siguiendo buenas prácticas en análisis de datos e inferencia causal, y estimando el efecto causal del tratamiento mediante la técnica de Double Machine Learning, con ayuda del paquete DoubleML en R.

Tipología: Estudio de casos, teóricos o prácticos, relacionados con la temática del Grado.

Objetivos planteados:

- Estudio de los fundamentos matemáticos en el campo de la Inferencia Causal.
- Estudio del Double Machine Learning como técnica de Inferencia Causal, principal objetivo de este trabajo.
- Estudio de la relación entre Double Machine Learning y el teorema de Frish-Waugh-Lovell.
- Generación de datos sintéticos a partir de grafos acíclicos dirigidos (DAG por sus siglas en inglés) para comprender e ilustrar el uso del Double Machine Learning.
- Aplicación de esta técnica a un conjunto de datos reales.

Bibliografía básica:

- [1] Neal, B. (2020). Introduction to Causal Inference from a Machine Learning Perspective.
- [2] Pearl, J., Glymour, M., & Jewell, N. P. (2016). Causal inference in statistics: A primer. John Wiley & Sons.
- [3] Rosenbaum, P. R. (2023). Causal inference. MIT Press.
- [4] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters.
- [5] Frisch, R., & Waugh, F. V. (1933). Partial time regressions as compared with individual trends. *Econometrica: Journal of the Econometric Society*, 387-401.
- [6] Lovell, M. C. (1963). Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association*, 58(304), 993-1010.
- [7] Smith MJ, Mansournia MA, Maringe C, Zivich PN, Cole SR, Leyrat C, Belot A, Rachet B, Luque-Fernandez MA. Introduction to computational causal inference using reproducible Stata, R, and Python code: A tutorial. *Stat Med*. 2022 Jan 30;41(2):407-432. doi: 10.1002/sim.9234. Epub 2021 Oct 28. PMID: 34713468; PMCID: PMC11795351.

Recomendaciones y orientaciones para el estudiante:

- Haber cursado las siguientes asignaturas obligatorias del Grado en Matemáticas:
 - Estadística Descriptiva e Introducción a la Probabilidad
 - Probabilidad
 - Inferencia Estadística
- Revisar el manual [1] y los siguientes enlaces como punto introductorio a la Inferencia Causal:
<https://onlinelibrary.wiley.com/doi/10.1002/sim.9234>
<https://github.com/migariane/TutorialComputationalCausalInferenceEstimators>

Plazas: 1

2. DATOS DEL TUTOR/A:

Nombre y apellidos: GUSTAVO RIVAS GERVILLA

Ámbito de conocimiento/Departamento: ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

Correo electrónico: griger@ugr.es

3. COTUTOR/A DE LA UGR (en su caso):

Nombre y apellidos: MIGUEL ÁNGEL LUQUE FERNÁNDEZ

Ámbito de conocimiento/Departamento: ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

Correo electrónico: mluquefe@ugr.es

4. COTUTOR/A EXTERNO/A (en su caso):

Nombre y apellidos:

Correo electrónico:

Nombre de la empresa o institución:

Dirección postal:

Puesto del tutor en la empresa o institución:

Centro de convenio Externo:

5. DATOS DEL ESTUDIANTE:

Nombre y apellidos: JESUS RUIZ LOPEZ

Correo electrónico: jesusruizlopez@correo.ugr.es