

UNIVERSIDAD NACIONAL DEL SUR										1	7
BAHIA BLANCA										ARGENTINA	
DEPARTAMENTO DE MATEMÁTICA											
PROGRAMA DE: Inferencia individualizada y Machine Learning										Curso / Seminario	
										Posgrado	
HORAS						PROFESOR RESPONSABLE					
30 hs						Dr. Fernando TOHMÉ					
						Dr. Fernando DELBIANCO					
REQUISITOS PREVIOS											
Probabilidad y Estadística. Es deseable, aunque no indispensable, familiaridad con nociones de Machine Learning y/o Econometría.											
DESCRIPCIÓN											
La primera clase introduce el cambio de paradigma desde la predicción global hacia la inferencia individualizada y transicional, destacando el problema de aprender sobre casos con tamaño muestral directo nulo. La segunda clase incorpora Predicción Conformal como herramienta libre de distribución para cuantificar incertidumbre en muestras finitas, incluso cuando el predictor subyacente es una caja negra. La tercera clase se concentra en algoritmos locales, métricas de similitud y controles sintéticos individualizados, con especial atención al problema de cold-start. La cuarta clase introduce herramientas de Teoría de Juegos cooperativos para explicar y auditar la contribución de los datos y variables a las predicciones, incluyendo el valor de Shapley, el nucléolo y los bargaining sets. La quinta clase aborda la combinación de fuentes heterogéneas mediante aprendizaje por fusión individualizada, distribuciones de confianza, criterios de Minimum Description Length (MDL) y aproximaciones bayesianas. La sexta clase cierra el recorrido con una perspectiva dinámica: trayectorias, cambios de régimen, análisis ordinal de patrones, entropía de permutación y límites intrínsecos del uso de los mismos datos para aprender y evaluar.											
El hilo conductor es metodológico y conceptual: construir inferencias útiles para unidades singulares sin abandonar exigencias de validez, transparencia, parsimonia, robustez y evaluación crítica. Al finalizar, los estudiantes podrán formular problemas de inferencia individualizada, seleccionar datos comparables, construir intervalos predictivos válidos, diseñar controles contrafácticos locales, atribuir importancia a la información utilizada, fusionar fuentes heterogéneas y reconocer los límites del aprendizaje automático aplicado.											
OBJETIVOS											
Objetivo general											
Brindar a los estudiantes un marco integrado para analizar, diseñar y evaluar métodos de aprendizaje automático orientados a inferencias individualizadas, en entornos caracterizados por heterogeneidad, incertidumbre, datos incompletos y necesidad de interpretabilidad.											
Objetivos específicos											
1. Distinguir entre inferencia poblacional, inferencia predictiva global e inferencia individualizada.											
2. Comprender el papel de la inferencia transicional cuando no existen observaciones directas de la unidad de interés.											
3. Utilizar Predicción Conformal para construir regiones o intervalos predictivos con garantías de cobertura en muestras finitas.											
AÑO	2026										

UNIVERSIDAD NACIONAL DEL SUR		2	7
BAHIA BLANCA		ARGENTINA	
DEPARTAMENTO DE MATEMÁTICA			
PROGRAMA DE: Inferencia individualizada y Machine Learning		Curso / Seminario	
		Posgrado	

4. Diseñar controles sintéticos individualizados.
5. Analizar la relevancia marginal de variables, observaciones y fuentes de información mediante herramientas de teoría de juegos cooperativos.
6. Integrar información heterogénea mediante esquemas de aprendizaje por fusión, distribuciones de confianza, MDL y criterios bayesianos.
7. Evaluar efectos individualizados en el tiempo, detectar cambios de régimen y reconocer límites fundamentales de evaluación y aprendizaje con los mismos datos.

MOTIVACIÓN O FUNDAMENTACIÓN DEL CURSO

El curso estudia el problema de la individualización en contextos de aprendizaje automático: cómo producir inferencias, predicciones y explicaciones confiables cuando el interés no se centra en la entidad promedio, sino en unidades, consultas o escenarios particulares. El punto de partida es la tensión entre los modelos globales, usualmente diseñados para capturar regularidades poblacionales, y las preguntas aplicadas que exigen respuestas localizadas: qué ocurrirá con una firma, un hogar, una región, un beneficiario de política pública o una unidad sin historial observable.

MECANISMO DE EVALUACIÓN

Se evaluará un trabajo final y su presentación oral. El trabajo deberá consistir en una nota técnica de 8 a 12 páginas, conteniendo:

1. Una pregunta de inferencia individualizada claramente formulada.
2. Descripción de la unidad objetivo y de las fuentes de información disponibles.
3. Justificación de la estrategia de comparación, conformalización, control sintético, atribución, fusión o análisis dinámico elegida.
4. Discusión de garantías, supuestos, riesgos de extrapolación y limitaciones.
5. Breve reflexión sobre interpretabilidad y uso de la inferencia resultante

PROGRAMA

Clase 1. Del modelo global a la inferencia individualizada

Motivación

El aprendizaje automático suele comenzar por modelos que resumen patrones poblacionales. Sin embargo, muchas decisiones aplicadas requieren responder preguntas sobre unidades específicas: un hogar, una firma, una escuela, un paciente, una jurisdicción o una política localizada.

AÑO	2026							
-----	------	--	--	--	--	--	--	--

UNIVERSIDAD NACIONAL DEL SUR							3	7
BAHIA BLANCA							ARGENTINA	
DEPARTAMENTO DE MATEMÁTICA								
PROGRAMA DE: Inferencia individualizada y Machine Learning						Curso / Seminario		
						Posgrado		
La clase introduce el problema conceptual que aparece cuando el tamaño muestral directo para la unidad de interés es nulo o insuficiente, y muestra por qué la extrapolación desde el promedio puede ser inadecuada en contextos de heterogeneidad.								
Contenidos								
Modelos predictivos globales, inferencia poblacional e inferencia para casos particulares. Heterogeneidad económica y límites de la aproximación asintótica clásica. El problema $n = 0$: consultas sobre unidades sin historial directo. Inferencia transicional: transferencia de información desde unidades similares. Criterios iniciales para decidir qué significa que dos unidades sean comparables.								
Clase 2. Predicción Conformal								
Motivación								
Los modelos de aprendizaje automático pueden alcanzar alto desempeño predictivo, pero frecuentemente operan como cajas negras y carecen de garantías claras de incertidumbre. La Predicción Conformal permite construir intervalos o conjuntos de predicción con garantías de cobertura en muestras finitas bajo supuestos débiles de intercambiabilidad. En esta clase se enfatiza su utilidad para decisiones bajo riesgo y su capacidad de funcionar con modelos agnósticos o mal especificados.								
Contenidos								
Problema de cuantificación de incertidumbre en modelos de caja negra. Predicción Conformal: intuición, puntajes de no conformidad y cobertura. Validez libre de distribución y garantías en muestras finitas. Conformalización de modelos arbitrarios de regresión y clasificación. Resiliencia de la cobertura frente a errores de especificación. Implicancias para decisiones bajo riesgo.								
Clase 3. Vecindades estadísticas, controles sintéticos individualizados y cold-start								
Motivación								
La inferencia individualizada depende de saber con qué unidades comparar el caso de interés. La clase estudia el trade-off entre relevancia y robustez: cuanto más estrecha es la vecindad local, más relevante puede ser la información transferida, pero menor es la base estadística disponible.								
Este problema aparece con fuerza en sistemas de recomendación, asignación, evaluación de políticas y construcción de contrafácticos.								
Contenidos								
Métricas de similitud, distancia, vecindades estadísticas y observaciones análogas. Algoritmos de filtrado local para seleccionar submuestras relevantes. Controles sintéticos individualizados y construcción de contrafácticos locales. Procedimientos de validación, simulación y sensibilidad. Problema de cold-start: unidades sin historial previo en recomendación, asignación o evaluación.								
AÑO	2026							

UNIVERSIDAD NACIONAL DEL SUR										4	7
BAHIA BLANCA										ARGENTINA	
DEPARTAMENTO DE MATEMÁTICA											
PROGRAMA DE: Inferencia individualizada y Machine Learning								Curso / Seminario			
								Posgrado			

Clase 4. Explicabilidad, atribución y teoría de juegos cooperativos

Motivación

En aplicaciones, no basta con predecir: es necesario explicar por qué se obtiene una predicción, qué variables o datos la sostienen y cuán estable es esa explicación. La clase introduce herramientas de teoría de juegos cooperativos para valorar contribuciones marginales de variables, observaciones o fuentes de información. El objetivo es conectar interpretabilidad y estabilidad de modelos complejos.

Contenidos

Transparencia y rendición de cuentas en modelos predictivos complejos. Juegos cooperativos transferibles como lenguaje para atribuir contribuciones. Valor de Shapley y atribución de importancia en clasificación, regresión y predicción local. Núcleo y conjuntos de negociación como herramientas de estabilidad y auditoría. Aporte marginal de datos de control y sensibilidad de explicaciones.

Clase 5. Fusión individualizada, MDL y criterios bayesianos

Motivación

Los datos aplicados rara vez provienen de una única fuente homogénea. Un analista puede tener encuestas, registros administrativos, series temporales, datos transaccionales, información geográfica o mediciones cualitativas. La clase estudia cómo fusionar información heterogénea de manera local, evitando tanto la fragmentación excesiva como la imposición de un modelo global demasiado rígido.

Contenidos

Fuentes múltiples y heterogéneas de información. Aprendizaje por fusión individualizada: Fusion Learning, iFusion e iGroup. Distribuciones de confianza locales y síntesis de evidencia. Modelado de subgrupos dinámicos y consultas específicas. Mínima Longitud de Descripción (MDL) y parsimonia de modelos locales. Aproximaciones bayesianas y comparación con regresiones penalizadas.

Clase 6. Dinámica individualizada, DIOPA y límites del aprendizaje

Motivación

Las unidades económicas no son estáticas. Sus trayectorias pueden mostrar persistencia, rupturas, cambios de régimen y patrones ordinales que no se capturan adecuadamente con estimaciones puntuales fijas. La clase final introduce el análisis dinámico de efectos individualizados y examina una tensión fundamental: los mismos datos no pueden usarse ilimitadamente para optimizar el aprendizaje predictivo y, al mismo tiempo, evaluar sus errores con perfecta independencia.

Contenidos

Efectos individualizados dinámicos y heterogeneidad temporal. Estimaciones locales móviles ponderadas. Algoritmo DIOPA: Dynamic Individualized Ordinal Pattern Analysis. Patrones ordinales, análisis simbólico de trayectorias y entropía de permutación. Detección de cambios de régimen estructural a nivel micro. Variante del

AÑO	2026							
-----	------	--	--	--	--	--	--	--

UNIVERSIDAD NACIONAL DEL SUR		5	7
BAHIA BLANCA		ARGENTINA	
DEPARTAMENTO DE MATEMÁTICA			
PROGRAMA DE: Inferencia individualizada y Machine Learning		Curso / Seminario	
		Posgrado	

principio de incertidumbre para aprendizaje automático: tensión entre optimización predictiva y evaluación de errores con los mismos datos. Perspectivas futuras: individualización y efectos de tratamiento heterogéneos, dinámicos y potencialmente no acotados

AÑO	2026							
-----	------	--	--	--	--	--	--	--

UNIVERSIDAD NACIONAL DEL SUR		6	7
BAHIA BLANCA		ARGENTINA	
DEPARTAMENTO DE MATEMÁTICA			
PROGRAMA DE: Inferencia individualizada y Machine Learning		Curso / Seminario	
		Posgrado	

BIBLIOGRAFIA

[1] Abadie, A., Diamond, A. y Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program. *Journal of the American Statistical Association*, 105(490), 493–505.

[2] Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2), 391–425.

[3] Angelopoulos, A. N. y Bates, S. (2023). Conformal prediction: A gentle introduction. *Foundations and Trends in Machine Learning*, 16(4), 494–591.

[4] Bandt, C. y Pompe, B. (2002). Permutation entropy: A natural complexity measure for time series. *Physical Review Letters*, 88(17), 174102.

[5] Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, 16(3), 199–231.

[6] Chernozhukov, V., Wüthrich, K. y Zhu, Y. (2018). Exact and robust conformal inference methods for predictive machine learning with dependent data. *Proceedings of Machine Learning Research*, 75, 732–749.

[7] Delbianco, F., Fioriti, A., and Tohmé, F. (2025) An assessment tool for academic research evaluation. *International Journal of Computational Economics and Econometrics*, 15(4), 392-420.

[8] Delbianco, F., Fioriti, A., Hernandez-Chanto, A., and Tohmé, F. (2020) A Markov-switching approach to the study of citations in academic journals. *Journal of Informetrics*, 14(4), 101081.

[9] Delbianco, F. and Tohmé, F. (2026). Individualized inference with relevant and superrelevant synthetic controls. *Manuscript*.

[10] Hannig, J., Iyer, H., Lai, R. C. S. y Lee, T. C. M. (2016). Generalized fiducial inference: A review and new results. *Journal of the American Statistical Association*, 111(515), 1346–1361.

[11] Hastie, T., Tibshirani, R. y Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

[12] Imbens, G. W. y Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.

[13] Molnar, C. (2022). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 2da edición. Disponible en línea.

[14] Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14(5), 465–471.

[15] Rissanen, J. (2007). *Information and Complexity in Statistical Modeling*. Springer.

UNIVERSIDAD NACIONAL DEL SUR		7	7
BAHIA BLANCA		ARGENTINA	
DEPARTAMENTO DE MATEMÁTICA			
PROGRAMA DE: Inferencia individualizada y Machine Learning		Curso / Seminario	
		Posgrado	

[16] Shafer, G. y Vovk, V. (2008). A tutorial on conformal prediction. Journal of Machine Learning Research, 9, 371–421. [17] Shapley, L. S. (1953). A value for n-person games. En H. W. Kuhn y A. W. Tucker (eds.), Contributions to the Theory of Games II, pp. 307–317. Princeton University Press. [18] Vapnik, V. N. (1998). Statistical Learning Theory. Wiley. [19] Vovk, V., Gammerman, A. y Shafer, G. (2005). Algorithmic Learning in a Random World. Springer. [20] Xie, M. y Singh, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter: A review. International Statistical Review, 81(1), 3–39		
AÑO	PROFESORES RESPONSABLES	DIRECTOR DEPARTAMENTO
2026	Fernando TOHMÉ Fernando DELBIANCO	

AÑO	2026							
------------	------	--	--	--	--	--	--	--