

SÍLABO 2024-2

I. INFORMACIÓN GENERAL

Nombre del curso	: Introducción a Machine Learning para Ciencias Sociales
Código del curso	: --
Carácter	: Curso de Formación continua
Créditos	: 2
Número de horas de teoría	: 3
Número de horas de práctica	: -
Profesor del curso	: Pavel Coronado Castellanos
Correo electrónico pucp	: pavel.coronado@pucp.edu.pe
Horario clases	: Sábado y Domingo 8:30am- 11:30am
Jefe de prácticas	: Josué Caldas, josue.caldas@pucp.edu.pe

II. SUMILLA

Machine Learning o Aprendizaje de Máquinas es una rama de la estadística y la inteligencia artificial que ha cobrado reciente importancia debido a su amplia gama de aplicaciones. Los métodos de Machine Learning (ML) son considerados como parte de la inteligencia artificial porque permiten realizar tareas que normalmente son hechas por humanos, como reconocer objetos en imágenes, detectar cáncer, interpretar el contenido o tono de un texto, entre otras tareas. En ciencias sociales, machine learning ha sido usado para predecir precios de residencias, decisiones de jueces, corrupción de gobiernos locales, migración de la población afro-americana, etc.

III. RESULTADOS DE APRENDIZAJE

Este curso proporcionará herramientas conceptuales, herramientas de implementación, pero también aplicaciones en temas de ciencias sociales. Todo esto para que los estudiantes puedan identificar rápidamente situaciones en las cuales los métodos de machine learning pueden ser aplicados en sus proyectos de tesis u otras investigaciones.

IV. CONTENIDO DEL CURSO

Bibliografía Requerida

- Gareth James, Daniela Witten, Trevor Hastie and Robert Tibsharani (2021), **An Introduction to Statistical Learning: with Applications in Python**. Springer. Second Edition. Intermediate level. (GWHT)

- Deisenroth Marc, Aldo Faisal, Cheng Soon Ong (2020) **Mathematics for Machine Learning**. Cambridge University. (DFS)
- Papers seleccionados por tema.

Introducción General a la Inteligencia Artificial. Econometría y Machine Learning. Predicción Vs. Causalidad. Repaso de Herramientas Estadísticas y de Optimización.

En esta sección se presenta de manera general las ideas base de la inteligencia artificial y el enfoque de machine learning. Luego de una revisión de las principales herramientas estadísticas y de optimización a ser usadas, se comparará el enfoque de ML con el enfoque cuantitativo tradicionalmente usado en ciencias sociales que enfatiza la obtención de estimadores causales (insegados). Machine Learning normalmente usa como indicador de aprendizaje la reducción del error de predicción, lo cual implica que sus herramientas se enfocan en la predicción.

- DFS Cap. 5.1-5.3
- GWHT Cap. 1 y 2
- Athey, Susan and Guido Imbens (2019) Machine Learning Methods that Economist should know about
- Breiman, Leo. (2001) “Statistical Modeling: The two cultures”. Statistical Science. Vol. 16 199-231.

Supervised Learning: Modelos Lineales. Conceptos Básicos. Capacidad óptima de los modelos; underfitting and overfitting; regularización, cross-validation, supervised/unsupervised learning. Regresiones Ridge y Lasso

En esta sección se introduce el concepto de complejidad de los modelos y como el exceso de complejidad en un modelo puede llevar al problema de overfitting. Luego se introduce el uso de las técnicas de regularización para controlar este problema. Finalmente, se presentan las herramientas de regresión que hacen uso de regularización (regresión Lasso y Ridge) y validación cruzada (cross validation).

- GWHT Cap. 3, 5 y 6
- Derenoncourt, Ellora (2020) “Can you move to opportunity? Evidence from The Great Migration” Working Paper.
- Mullainathan, Sendhil y Jann Spies (2017). Machine Learning: An Applied Econometric Approach. Journal of Economic Perspectives. Vol. 31. Number 2.
- Ward, Michael, Brian Greenhill, Kristin Bakee (2010). The perils of policy by p-value: Predicting civil conflicts. Journal of Peace Research. Vol. 47

Supervised Learning: Nonlinear Models Methods: Trees, Boosting, Bootstrap

En esta sección se presentan las herramientas típicas de machine learning para tareas de clasificación. Classification and regression Trees (Breiman, 1984). También se presentarán

las técnicas de Ensemble que optimizan la predicción de estos algoritmos (Breiman, 2001; Friedman, 2001).

- GWHT Cap. 8
- Ash, Elliot; Sergio Galleta; and Tommaso Giommoni (2020) “A Machine Learning Approach to Analyze and Support Anti-Corruption Policy” Working Paper
- Raval, D.; Rosenbaum, T., & Wilson, N. E. (2019). “How Do Machine Learning Algorithms Perform in Changing Environments? Evidence from Disaster Induced Hospital Closures.”
- Kleinberg, Jon; Himabindu Lakkaraju, Jure Leskive, Sendhil Mullainathan (2017) “Human Decisions and Machine Predictions” NBER Working Papers. N. 23180

Introducción a Redes Neuronales y Deep Learning

En esta sección se presentan los fundamentos de redes neuronales. Comenzando por redes neuronales de una capa (single layer), multicapas, convolucionales, entre otras.

- GWHT Cap. 10
- Arman Khachiyani, Anthony Thomas, Huye Zhou, Gordon H. Hanson, Alex Cloninger, Tajana Rosing, and Amit Khandelwal (2021) Using Neural Networks to Predict Micro-Spatial Economic Growth. NBER Working Paper No. 29569

V. METODOLOGÍA

El contenido del curso se desarrollará mediante sesiones sincrónicas. El alumno aprenderá el contenido del curso, también, mediante el desarrollo de ejercicios y laboratorios calificados. Finalmente, el alumnos plasmará los conocimientos aprendidos en un proyecto de investigación.

VI. EVALUACIÓN

(La forma de evaluación, exámenes, controles, ejercicios serán aquí detalladas con sus pesos y se explicará también si la evaluación ocurrirá de forma sincrónica o asincrónica. Se recomienda colocar aquí las pautas de todas las formas de evaluación, de no ser así deberán enviarse a los estudiantes con anterioridad a cada evaluación. El acta de calificación del curso se elabora en base a la fórmula de, por lo que no puede modificarse).

Nº	Tipo de Evaluación	Ponderación sobre la nota final
	Lista de Ejercicios y Laboratorios (LE) en grupos de 4	30%
	Presentación de idea General para el Proyecto Final (Una cara) 22 de Octubre en grupos de 2 ó 3 máximo	10%
	Exposición en Clase 2-9 Noviembre	15%
	Propuesta de Proyecto (PP) – 14 de Noviembre	35%
	Participación Continua (PC)	10%

Fórmula de calificación: Nota Final =0.3*LE+0.6*PP+0.1*PC

VII. BIBLIOGRAFÍA OBLIGATORIA¹

Gareth James, Daniela Witten, Trevor Hastie and Robert Tibsharani (2021), **An Introduction to Statistical Learning: with Applications in R**. Springer. Second Edition. Intermediate level. (GWHT). Puede acceder al libro [aquí](#).

Deisenroth Marc, Aldo Faisal, Cheng Soon Ong (2020) **Mathematics for Machine Learning**. Cambridge University Press. (DFS)

VIII. CRONOGRAMA

(En esta sección se elabora un cronograma con fecha y tema de cada sesión agregando para el semestre de oferta de cursos a distancia una breve referencia a cómo se trabajará en la semana e.j. sesión sincrónica y ejercicio; sesión sincrónica y participación en foro; sesión sincrónica y autoestudio con entrega de reporte de lectura)

Clase	TEMA/CONTENIDOS	Práctica Dirigida
28 de Septiembre	Introducción al curso e Introducción General a la Inteligencia Artificial. Econometría y Machine Learning. Predicción Vs. Causalidad. Repaso Est. y optimización.	
29 de Septiembre	Repaso de Herramientas Estadísticas y de Optimización. 2	
5 de octubre	No Clases Combate de Angamos	Laboratorio Dirigido 1. (2 de Octubre)
6 de Octubre	No Clases Combate de Angamos	
12 de Octubre	Modelos Lineales. Conceptos Básicos. Capacidad óptima de los modelos; underfitting and overfitting; regularización, cross-validation,	(9 de Octubre).Laboratorio Dirigido 2: Aplicación en Python, Simulaciones- Modelos Lineales. Entrega PC1
13 de Octubre	Métodos de Remuestreo. Crossvalidation. Bootstrap. Supervised learning. Regresiones Ridge y Lasso. Parte 1	
19 de Octubre	Supervised learning. Regresiones Ridge y Lasso. Parte 2	Laboratorio dirigido 3 (16 de Octubre). Aplicación en Python: Simulaciones -Modelos Lineales

¹ Recordar que la bibliografía obligatoria debe estar disponible en el PAIDEIA del curso.

20 de Octubre	Supervised Learning: Nonlinear Models Methods: Trees, Boosting, Bootstrap Parte 1	Entrega: Ideas Generales Entrega PC2. Devolución PC1
26 de Octubre	Introducción a Deep learning: Redes Neuronales Multicapa. Redes Neuronales convolutivas - Recurrentes	Laboratorio dirigido 4 (23 de Octubre) .Aplicación en Python: Árboles de decisión
27 de Octubre	Introducción a Deep learning: Redes Neuronales Multicapa. Redes Neuronales convolutivas - Recurrentes	
2 de Noviembre	Presentaciones	Laboratorio dirigido 5 (30 de Octubre) Aplicación en Python: Redes neuronales. Entrega PC3. Devolución PC2
3 de Noviembre	Presentaciones	
9 de Noviembre	Presentaciones	Devolución PC3 (6 de Noviembre)

La evaluación de todos los trabajos contemplará el respeto de los derechos de autor. En este marco, cualquier indicio de plagio tendrá como consecuencia la nota cero. Esta medida es independiente del proceso disciplinario que la Secretaría Académica de la facultad estime iniciar según cada caso. Para obtener más información sobre el citado visitar el siguiente sitio web: www.pucp.edu.pe/documento/pucp/plagio.pdf