

Planeación General del Curso

Inteligencia de Datos		
Universidad Iberoamericana Ciudad de México		
Programa: Actuaría	Semestre ideal: Octavo semestre	
Conocimientos esenciales: Cálculo de varias variables, optimización no lineal, álgebra, álgebra lineal, probabilidad, programación estructurada (Python).		
Profesor: Luis Norberto Zúñiga Morales	Grupo: B	
Fines de aprendizaje: <ol style="list-style-type: none">1. Estudiar y analizar el planteamiento de modelos de Aprendizaje Automático más complejos ampliamente utilizados en la literatura y la industria.2. Implementar los modelos estudiados utilizando diversos conjuntos de datos de repositorios, interpretar los resultados obtenidos y mejorar su desempeño mediante su análisis.3. Aprender a usar diversas herramientas de Python para la implementación de los modelos de Aprendizaje Automático vistos en clase.4. Analizar y comprender la razón por la que ciertos modelos de Aprendizaje Automático se utilizan en ciertas aplicaciones.5. Comprender el ciclo de vida de un proyecto de datos para aplicarlo con un problema simulacro.6. Estudiar diversos patrones de diseño que se utilizan al momento de entrenar modelos de Machine Learning.		
Objetivo: Analizar las tendencias y paradigmas de la Ciencia de Datos, contemplando el fundamento matemático y desarrollo de distintos modelos para su implementación en diversos proyectos para comprender el ciclo de vida de un proyecto de Ciencia de Datos.		
Semanas: 17	Horas: 4	Total de horas: 68

Temario

1. Máquinas de Vectores de Soporte
 - a. Problema de clasificación lineal
 - b. Problema del margen suave
 - c. Problema de clasificación no lineal
 - d. Clasificación multiclase
 - e. Ejemplos de aplicación
2. Árboles de Decisión
 - a. Caso de clasificación

- b. Caso de regresión
 - c. Ejemplos de aplicación
3. Ensemble Learning
 - a. Voting classifiers
 - b. Bagging y Pasting
 - c. Boosting: AdaBoost
 - d. Bosques aleatorios
 - e. Ejemplos de aplicación
4. Ciclo de vida de un proyecto de ciencia de datos
5. Patrones de diseño en Machine Learning
 - a. Patrones de representación de datos
 - b. Patrones de diseño de representación de problemas
 - c. Patrones de entrenamiento de modelos
6. Simulacro de un proyecto de Machine Learning
 - a. Repositorios de datos y APIs
 - b. Entrenamiento de modelos aplicados a problemas
 - c. Presentación de resultados

Actividades Propuestas

1. **Presentación de los temas** contemplados en el curso por parte del profesor.
2. **Prácticas de laboratorio de cómputo** mediante [Google Colab](#) para implementar distintos algoritmos de Ciencia de Datos. Los conjuntos de datos propuestos para cada práctica se pueden obtener de los repositorios discutidos en clase o se pueden construir por medio de APIs ([Twitter API](#), [NYT API](#), etc.) o *web scraping*.
3. **Lecturas de artículos científicos** relativos a aplicaciones, paradigmas y filosofía de la Ciencia de Datos. Dichas lecturas pueden ser útiles como una introducción o punto de partida para ejemplificar el objetivo de las prácticas de laboratorio, y para entender el panorama actual de la Ciencia de Datos como una disciplina en la academia y/o la industria.
4. **Evaluaciones** para determinar el avance teórico y práctico del alumno.
5. **Proyecto final** para simular todo el proceso que se lleva a cabo en un proyecto de Machine Learning en la vida real.
6. **Creación de una página de Wikipedia** en español sobre un tema relativo al Machine Learning y la Ciencia de Datos. El tema se deja a elección del estudiante. Proyecto para el final del semestre que se evalúa en tres partes: dos borradores para medir el avance del alumno y la entrega final ya publicada en el sitio de Wikipedia.

Fechas Importantes

- **Primer examen:** 4 de marzo de 2024
- **Segundo examen:** 10 de abril de 2024
- **Proyecto final:** hasta el 6 de mayo de 2024.
- **Proyecto wikipedia (primera entrega):** hasta el 20 de marzo de 2024.
- **Proyecto wikipedia (segunda entrega):** hasta el 24 de abril de 2024.
- **Proyecto wikipedia (entrega final):** hasta el 6 de mayo de 2024.

- **Entrega de tareas:** hasta el 6 de mayo de 2024.
- **Entrega de calificaciones:** 8 de mayo de 2024.
- **Fin de clases:** 11 de mayo de 2024 (8 de mayo

Bibliografía

1. Abu-Mostafa, Y. S., Magdon-Ismail, M., & Lin, H.-T. (2012). *Learning from Data: A Short Course*. AMLBook.com.
2. Calin, O. (2020). *Deep Learning Architectures: A Mathematical Approach*. Springer International Publishing.
3. Courville, A., Bengio, Y., & Goodfellow, I. (2016). *Deep Learning*. MIT Press.
4. Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media, Inc.
5. Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
6. Lakshmanan, Valliappa, et al. *Machine Learning Design Patterns: Solutions to Common Challenges in Data Preparation, Model Building, and MLOps*. O'Reilly, 2020.
7. Müller, A. C., & Guido, S. (2016). *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media, Inc.

Instrumentos de Evaluación

Instrumento	Porcentaje
Tareas y prácticas de cómputo	20%
Proyecto Wikipedia	20% (5%-5%-10%)
Primera evaluación (examen)	20%
Segunda evaluación (examen)	20%
Tercera evaluación (proyecto final)	20%

Total	100%
--------------	------

Cronograma

Módulo	Fecha	Tema	Actividad
Introducción al curso	08/01/24	Bienvenida	Presentación del Curso
			Syllabus
			Preguntas sobre el curso
	10/01/24	Repaso de Machine Learning	¿Qué es el Machine Learning?
			Componentes del aprendizaje
			Tipos de aprendizaje
			Intro al perceptrón
Máquinas de Vectores de Soporte	15/01/24	Caso lineal	Planteamiento del problema
			Caso del margen duro
			Problema de optimización
			Estimación de parámetros del modelo
	17/01/24	Margen suave	Planteamiento del problema
			Caso del marge suave
			Problema de optimización
			Estimación de parámetros del modelo
	22/01/24	Caso no lineal	Planteamiento del problema
			Caso del margen duro
			Truco del Kernel
			Problema de optimización
			Estimación de parámetros del modelo
		Clasificación multiclase	Presentación del modelo
			Criterios de clasificación multiclase
			Estimación de parámetros del modelo
			Actividad: Infografía MVS

	24/01/24	Aplicación de Máquinas de Vectores de Soporte	Práctica 1: Máquinas de Vectores de Soporte en Scikit-Learn Tarea 1: Ejercicios notas MVS
Árboles de Decisión	29/01/24	Árboles de decisión	Presentación del modelo
			Problema de clasificación
			Estimación de parámetros del modelo
	31/01/24		Práctica 2: Árboles de decisión en Scikit-Learn Tarea 2: Árboles de decisión para regresión
	05/02/24	No hay clase	
Ensemble Learning	07/02/24	Voting Classifiers	Introducción al Ensemble Learning
			Voting Classifiers
			Práctica 3: Voting Classifiers en Scikit-Learn
	12/02/24	Bagging y Pasting	Bagging y Pasting
			Práctica 4: Bagging en Scikit-Learn
	14/02/24	Boosting	Idea general del Boosting
			AdaBoost
			Práctica 5: Boosting en Scikit-Learn
	19/02/24	Bosques Aleatorios	Idea general de Bosques Aleatorios
			Práctica 6: Bosques Aleatorios en Scikit-Learn
	21/02/24		Varianza y Ensemble Learning
			Actividad: Infografía Ensemble Learning
Tarea 3: Notas Ensemble Learning			
26/02/24	Ejemplos de aplicación	Práctica 7: Ensemble Battle Royale	
28/02/24		Práctica 8: Predicción de Demanda Energética	

Evaluación	04/03/24	Primer Examen		
Patrones de diseño en Machine Learning	06/03/24	Ciclo de vida de un proyecto de ciencia de datos	Componentes del ciclo de vida de un proyecto de ciencia de datos	
	11/03/24	Patrones de Diseño en Machine Learning	¿Qué es un patrón de diseño?	
			Patrones de representación de datos	
			Práctica 9: Patrones de representación de datos	
	13/03/24		No hay clase	
	18/03/24		Patrones de representación de problemas	
	20/03/24		Entrega del primer borrador proyecto Wikipedia	
			Práctica 10: Patrones de representación de problemas	
	01/04/24		Patrones de entrenamiento de modelos de Machine Learning	
03/04/24	Práctica 11: Patrones de entrenamiento de modelos de Machine Learning			
08/04/24				
Evaluación	10/04/24	Segundo Examen		
Simulacro de un proyecto de Machine Learning	15/04/24	Repositorios de datos y APIs	APIs para obtener información: New York Times	
			Repositorios de datos en internet	
			Actividad: Manejo de la API del NYT	
	17/04/24	Proyecto: Análisis de Sentimientos en Twitter	Introducción al análisis de sentimientos	
			Modelos de lenguaje: bolsa de palabras y embeddings (fastText)	
	22/04/24		Práctica 12a: Análisis de sentimientos aplicado a tweets en inglés	
	24/04/24		Práctica 12b: Optimización de parámetros	
Entrega del segundo borrador proyecto Wikipedia.				
Asesorías	29/04/24		Asesorías para el proyecto final	

	01/05/24	No hay clase
Entrega de Calificaciones	06/05/24	Entrega de proyecto final, tareas y prácticas (extemporáneas) para su evaluación
	08/05/24	Entrega y corrección de calificaciones
Fin del curso		