

Planeación General del Curso

Inteligencia de Datos		
Profesor: Luis Norberto Zuñiga Morales		Grupo: A
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. Aprender a usar diversas herramientas para la recolección y construcción de un conjunto de datos.5. Analizar y comprender la razón por la que ciertos modelos de Aprendizaje Automático se utilizan en ciertas aplicaciones.6. Comprender el ciclo de vida de un proyecto de datos para aplicarlo con un conjunto de datos.		
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
 - i. Problema linealmente separable
 - ii. Margen Suave
 - b. Problema de Clasificación no Lineal
 - i. Truco del Kernel
 - c. Clasificación Multiclase
 - d. Ejemplos de aplicación
2. Bias-Variance Tradeoff
3. Ensemble Learning
 - a. Voting Classifiers
 - b. Bagging
 - c. Pasting
 - d. Random Forests

- e. Boosting
 - i. AdaBoosting
 - ii. Gradient Boosting
 - f. Ejemplos de aplicación
- 4. Visualización de Datos
- 5. Detección de Anomalías
- 6. Ciclo de vida de un proyecto de ciencia de datos
 - a. Análisis del campo de estudio
 - b. Recolección de datos
 - c. Preparación de los datos
 - d. Análisis exploratorio
 - e. Selección y entrenamiento del modelo
 - f. Evaluación del modelo
 - g. Ajuste del modelo
 - h. Despliegue del modelo
 - i. Ejemplos de aplicación
- 7. Simulacro de un proyecto de ciencia de datos
- 8. Entrenamiento de redes neuronales
 - a. Introducción a redes neuronales en Keras
 - b. Problemas con redes neuronales
 - i. Desvanecimiento del gradiente
 - ii. Transfer Learning
 - iii. Entrenamiento de una red
 - iv. Sobreajuste y regularización

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. **Proyecto final** para simular todo el proceso que se lleva a cabo en un proyecto de ciencia de datos en la vida real.
5. **Creación de una página de Wikipedia** en español sobre un tema relativo a la Ciencia de Datos. El tema se deja a elección del estudiante. Proyecto para el final del semestre.

Fechas Importantes

- **Fin de clases:** 10 de diciembre 2022 (7 de diciembre)

- **Registro de notas:** hasta el 10 de diciembre 2022
- **Trabajo final:** Semana del 28 de noviembre al 2 de diciembre 2022
- **Entrega de tareas:** Revisión siguiente clase; entrega del documento final semana del 28 de noviembre al 2 de diciembre 2022.

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. Dixon, M. F., Halperin, I., & Bilokon, P. (2020). *Machine Learning in Finance: From Theory to Practice*. Springer International Publishing.
5. Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media, Inc.
6. Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
7. Lopez de Prado, M. (2018). *Advances in Financial Machine Learning*. Wiley.
8. 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	35%
Proyecto Wikipedia	25%
Proyecto final	40%
Evaluaciones Adicionales	10%

Total	110%
--------------	------