

INTRODUCCIÓN A LA CIENCIA DE DATOS

NOMBRE DEL CURSO: Introducción a la Ciencia de Datos

PRERREQUISITO: FISI 2028 (Métodos Computacionales) / (Métodos computacionales 2)

CRÉDITOS: 3 créditos pregrado. 4 créditos posgrado.

CÓDIGO DEL CURSO: FISI 3915 - FISI 4915

UNIDAD ACADÉMICA: Departamento de Física

PERIODO ACADÉMICO: 202210

HORARIO: Mi-Vi, 15:30 -16:45

NOMBRE PROFESOR MAGISTRAL: Jaime Ernesto Forero Romero

CORREO ELECTRÓNICO: je.forero@uniandes.edu.co

HORARIO Y LUGAR DE ATENCIÓN: con cita previa.

NOMBRE PROFESOR LABORATORIO: John Fredy Suárez Pérez

CORREO ELECTRÓNICO: jf.suarez@uniandes.edu.co

HORARIO DE ATENCIÓN: con cita previa.

I Introducción

La ciencia de datos (Data Science) se empieza a posicionar en el centro de todas las áreas técnicas y científicas, dentro y afuera del ámbito académico. El curso de *Introducción a la Ciencia de Datos* presenta un panorama general de los principios y técnicas computacionales básicas para una persona que desea iniciarse en la Ciencia de Datos. Para esto se propone profundizar sus conocimientos en dos áreas: estadística descriptiva algoritmos para extraer patrones en conjuntos de datos. Se asume que los estudiantes de este curso ya tienen conocimientos básicos en métodos computacionales equivalentes al nivel del curso Métodos Computacionales (FISI-2028) del antiguo pensum de la carrera de Física, o Métodos Computacionales 2 del nuevo pensum. El lenguaje de programación será Python.

II Objetivos

El objetivo principal del curso es presentar métodos y algoritmos para extraer conclusiones a partir de un conjunto de datos.

III Competencias a desarrollar

Al finalizar el curso, se espera que el estudiante esté en capacidad de:

- formular preguntas o hipótesis sobre las propiedades un conjunto de datos.
- responder preguntas o descartar hipótesis hechas sobre las propiedades un conjunto de datos.
- comunicar de manera clara las conclusiones de análisis hechos sobre un conjunto de datos.

IV Contenido por semanas

Semana 1. Presentación del curso. Repaso fundamentos de probabilidad. Repaso de estadística descriptiva. Variables aleatorias. Distribuciones.

Semana 2. Prueba de hipótesis. Bootstrapping.

Semana 3. Regresión lineal.

Semana 4. Presentaciones de los resultados del proyecto #1
Semana 5. Cadenas de Markov.
Semana 6. Teorema de Bayes. Estimación de parámetros bayesiana.
Semana 7. Regresión logística.
Semana 8. Presentaciones de los resultados del proyecto #2
Semana 9. Árboles de decisión.
Semana 10 Bosques aleatorios.
Semana 11. Gradient boosting.
Semana 12. Presentaciones de los resultados del proyecto #3
Semana 13. Perceptrón
Semana 14. Redes neuronales.
Semana 15. Redes Neuronales Convolucionales.
Semana 16. Presentaciones de los resultados del proyecto #4

El repositorio del curso es: <https://github.com/ComputoCienciasUniandes/IntroDataScience>.

V Metodología

En las sesiones de los miércoles se hará énfasis en los conceptos teóricos. En las sesiones de los viernes se hará énfasis en la práctica computacional. Es necesario que los estudiantes preparen antes de cada clase el tema correspondiente siguiendo las lecturas preparatorias recomendadas por SICUA.

VI Criterios de evaluación

Se evaluarán 4 Proyectos (25 % cada uno para estudiantes que inscribieron el curso de 3 créditos, 20 % cada uno para estudiantes que inscribieron el curso de 4 créditos). Cada cuatro semanas los estudiante deben presentar la solución a una pregunta con las herramientas vistas hasta el momento en el curso. Se evalúa la originalidad de la pregunta, la claridad del planteamiento de la pregunta, la relevancia del pregunta, la solidez de los experimentos computacionales para responderla y la claridad de la explicación de los resultados.

Para los estudiantes que inscribieron el curso de 4 créditos hay un 20 % que se califica a partir de la toma de apuntes. Cada semana habrá tres estudiantes encargados de tomar los apuntes de la clase para pasarlos a L^AT_EX en un overleaf del curso.

VII Bibliografía

Bibliografía principal:

- *Introduction to Probability and Statistics for Engineers and Scientists*. Sheldon M. Ross. Third Edition. Elsevier Academic Press. 2004.
- *Pattern Recognition and Machine Learning*. C. M. Bishop, Springer, 2006.
- *The Data Science Manual*. S. S. Skienna, Springer, 2017.
- *Deep Learning*, I. Goodfellow, Y. Bengio A. Courville, MIT Press 2016
- *A Comprehensive Guide to Machine Learning*, S. Nasiriany, G. Thomas, W. Wang, A. Yang, Berkeley, 2019
- *Python Data Science Handbook*. J. VanderPlas, O'Reilly, 2016.
- *An Introduction to Statistical Learning with Applications in R*, G. James, D. Witten, T. Hastie, R. Tibshirani, Springer, 2015
- *A Student's Guide to Numerical Methods*. I. H. Hutchinson. Cambdrige, 2015
- *Data Analysis: A Bayesian Tutorial*. D. S. Sivia, J. Skilling. Second Edition, Oxford Science Publications. 2006