

INTRODUCCIÓN A LA CIENCIA DE DATOS

NOMBRE DEL CURSO: Introducción a la Ciencia de Datos

PRERREQUISITO: FISI 2028 (Métodos Computacionales)

CRÉDITOS: 3 créditos pregrado. 4 créditos posgrado.

CÓDIGO DEL CURSO: FISI 3915 - FISI 4915

UNIDAD ACADÉMICA: Departamento de Física

PERIODO ACADÉMICO: 202010

HORARIO: Mi-Vi, 9:30AM-10:50AM

NOMBRE PROFESOR MAGISTRAL: Jaime Ernesto Forero Romero

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

HORARIO Y LUGAR DE ATENCIÓN: Lunes 15:00 a 16:00, Oficina Ip208.

I Introducción

La ciencia de datos (Data Science) se encuentra hoy en día en todas las áreas técnicas y científicas, dentro y afuera del ámbito académico. Esto se debe en gran parte a que la capacidad de procesar grandes cantidades de datos en computadoras de alto rendimiento ha disminuido en costo monetario y en complejidad.

El curso de *Introducción a la Ciencia de Datos* presenta estas posibilidades computacionales a estudiantes de diferentes disciplinas científicas. Para esto se propone profundizar sus conocimientos en dos áreas: métodos de descripción estadística de datos y la implementación de algoritmos para extraer patrones presentes en diferentes tipos 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). El lenguaje de programación será Python.

II Objetivos

El objetivo principal del curso es presentar métodos, proceso 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:

- hacer hipótesis sobre las propiedades un conjunto de datos.
- descartar hipótesis hechas sobre las propiedades un conjunto de datos.

IV Contenido por semanas

Semana 1. Presentación del curso. Relaciones entre ciencia de datos, machine learning e inteligencia artificial.

Semana 2. Repaso fundamentos de probabilidad. Repaso de estadística descriptiva. Variables aleatorias. Distribuciones.

Semana 3 Prueba de hipótesis. Bootstrapping.

Semana 4 Cadenas de Markov.

Semana 5 Teorema de Bayes. Estimación de parámetros bayesiana.

Semana 6 Regresión lineal.

Semana 7 Regresión logística.

Semana 8 Árboles de decisión.

Semana 9 Bosques aleatorios.

Semana 10 Gradient boosting.

Semana 11 Perceptrón. Redes neuronales.

Semana 12 Redes Neuronales Convolucionales.

Semana 13 De la Ciencia a la Ciencia de Datos. Charlas y testimonios.

Semana 14 Presentaciones con los enunciados de los retos. Elección de los retos y conformación de equipos.

Semana 15 Presentación de avance de soluciones de los retos.

Semana 16 Presentación final de la solución de los retos.

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

V Metodología

En las sesiones magistrales, luego de presentar un resumen de los conceptos teóricos, se hará énfasis en la práctica computacional. Para que esto funcione es necesario que los estudiantes estudien el tema correspondiente **antes de cada clase** siguiendo las lecturas preparatorias recomendadas por SICUA.

VI Criterios de evaluación

Las componentes que reciben calificación en la Magistral (en paréntesis su contribución a la nota definitiva) son las siguientes:

- 3 Ejercicios (25 % cada uno). Cada cuatro semanas se deja planteado un ejercicio para resolver. Solamente se evalúa el código y los resultados del código.
- Reto (25 %). El reto es un ejercicio abierto donde se da solución a una pregunta con las herramientas vistas en el curso. Se evalúa la claridad del planteamiento del problema, la relevancia del problema, la solidez de los experimentos computacionales para resolverlo y la claridad de la explicación de los resultados.

VII Bibliografia

Bibliografia principal:

- *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. Cambridge, 2015
- *Data Analysis: A Bayesian Tutorial*. D. S. Sivia, J. Skilling. Second Edition, Oxford Science Publications. 2006