

PROGRAMA DE CURSO

Código	Nombre			
No rellenar	ANÁLISIS PARA CIENCIA DE DATOS			
Nombre en Inglés				
ANALYSIS FOR DATA SCIENCE				
SCT	Horas semestrales	Horas de Cátedra	Horas de ayudantías y laboratorios	Horas de Trabajo Personal
9	270	72	34.5	163.5
Requisitos			Carácter del Curso	
- Cálculo Avanzado - Álgebra Lineal - Matemáticas Discretas - Probabilidades y Estadística (co-requisito)			Obligatorio de Carrera Ingeniería Civil en Modelamiento Matemático de datos	
Resultados de Aprendizaje				
Se espera que el alumno sea capaz de: ● Conocer las dificultades inherentes al trabajo en ciencia de datos y la importancia de los fundamentos matemáticos para poder abordarlas de manera exitosa. ● Aplicar los fundamentos matemáticos de Álgebra Lineal como proyección, mínimos cuadrados, técnicas espectrales y análisis de componentes principales en diversos contextos. ● Entender conceptos y herramientas de espacios de Hilbert. Aplicar estos conceptos en análisis armónico y en teoría de probabilidades. ● Entender los elementos y herramientas de teoría de probabilidad. Aplicar estos conceptos al diseño de algoritmos aleatorizados para datos masivos y de alta dimensión. ● Implementar las herramientas aprendidas dentro de cada unidad con datos reales, incluido el análisis crítico de sus resultados, en aplicaciones que incluyen por ejemplo: problemas de regresión, reducción de dimensionalidad, clusterización o procesamiento de señales e imágenes.				

Metodología Docente	Evaluación General
Clases expositivas	Controles
Ayudantías	Proyectos por unidad
Laboratorios de computación	Examen

Unidades Temáticas

Número	Nombre de la Unidad	Duración en Semanas
1	Introducción a la Ciencia de Datos	1
Contenidos		
● Ejemplos de casos en: ☐ Aprendizaje de Máquina. ☐ Estadística y Probabilidades. ☐ Optimización e investigación de operaciones.		

Número	Nombre de la Unidad	Duración en Semanas
2	Álgebra Lineal	4
Contenidos		
● Repaso de Espacios vectoriales en dimensión finita. ● Producto interno: Desigualdad de Cauchy-Schwarz, proyección en subespacios. ● Formulación de un problema de optimización: mínimos cuadrados y aplicaciones a aprendizaje de máquinas. Condiciones de optimalidad (sin restricciones y Lagrange). ● Inversas generalizadas. ● Normas matriciales y distancia. ● Sumas directas. ● Teorema espectral. ● Descomposición en valores singulares (representación de transformaciones lineales). ● Análisis de componentes principales y aplicaciones. ● Matrices dispersas (pre-condicionamiento).		

Número	Nombre de la Unidad	Duración en Semanas
3	Espacios de Hilbert	6
Contenidos		
● Definición y propiedades. ◦ Funcionales Lineales y Bilineales. ◦ Formas Cuadráticas. ◦ Productos Internos y Normas. Nociones básicas de topología.		

◦ Desigualdad de Cauchy-Schwarz y Bessel. ◦ Ortogonalidad y complementos ortogonales. ◦ Bases. ◦ Dimensión. ◦ Acotamiento de funcionales. ● Ejemplo: Espacio de funciones de cuadrado integrables (L_2). ● Convergencia fuerte y débil. ● Dualidad (Teorema de Riesz). ● Proyección sobre un convexo. ● Wavelets, Fourier y Transformada Rápida de Fourier: teoría y aplicaciones en procesamiento de imágenes y señales.
--

Número	Nombre de la Unidad	Duración en Semanas
4	Probabilidad	5
Contenidos		
● Introducción a la medibilidad y sigmas álgebras. ● Esperanza y Varianza. ● Independencia, correlación y covarianza. Esperanzas condicionales, independencia condicional (en L_2). ● Normales multivariadas. ● Nociones de convergencia en probabilidad. ● Ley de los grandes números y Teorema Central del Límite. ● Desigualdades de concentración: Markov, Chebyshev, Chernoff (Hoeffding). ● Reducción de dimensionalidad: Lemma de Johnson-Lindenstrauss. ● Aplicaciones: k-nearest neighbor, Sketching, Sample Complexity.		

Bibliografía General
● Sheldon Axler, "Linear Algebra Done Right", Third Edition, 2015. ● Lokenath Debnath and Piotr Mikusiński, "Hilbert Spaces with Applications", 2005. ● Alan Karr, "Probability", 1993.
Bibliografía Complementaria

- Avrim Blum, John Hopcroft, Ravindran Kannan, "Foundations of Data Science", 2018.
- Paul Halmos, "Introduction to Hilbert Space and the Theory of Spectral Multiplicity", 2013.
- Haim Brezis, "Functional Analysis, Sobolev Spaces and Partial Differential Equations", 2011.
- Pedro Morettin, "Wavelets in Functional Data Analysis", 2017.
- Jean-Pierre Florens, "Econometric Modeling and Inference", 2012.
- Yung Tong, "The Multivariate Normal Distribution", 1990
- Sanjoy Dasgupta, Anupam Gupta, "An elementary proof of a theorem of Johnson and Lindenstrauss". Random Structures & Algorithms, 22 (1): 60–65, 2003.

Vigencia desde:	2019
Elaborado por:	Sebastián Donoso, Pedro Pérez, Cristóbal Quiñinao, Ernesto San Martín, David Sossa, Víctor Verdugo, José Verschae, Emilio Vilches.