

	PROGRAMA OFICIAL DE CURSO (Pregrado y Posgrado)
	UNIVERSIDAD DE ANTIOQUIA

1. INFORMACIÓN GENERAL

Unidad Académica:	Facultad de Ciencias Exactas y Naturales		
Programa académico al que pertenece:	Pregrado en Física		
Programas académicos a los cuales se ofrece el curso:	Pregrado en Física, Pregrado en Astronomía		
Vigencia:	2021-1	Código curso:	SOLICITAR A ADMISIONES
Nombre del curso:	Aprendizaje Estadístico		
Área o componente de formación del currículo: Profundización Curso Profesional			
Tipo de curso:	Teórico	Créditos académicos¹:	4
Características del curso:	Validable <input type="checkbox"/>	Habilitable x	Clasificable <input type="checkbox"/> Evaluación de suficiencia <input type="checkbox"/>
Modalidad del curso: Virtual			
Pre-requisitos:	Fundamentos de Mecánica Cuántica (0302703).		
Co-requisitos:	No tiene.		
Horas docencia directa: 4		Horas de trabajo independiente :	8
Horas totales del curso: 12			
Profesor(a) que elaboró: Sebastián Duque Mesa		Correo electrónico:	sebastian.duquem@udea.edu.co

2. INFORMACIÓN ESPECÍFICA

Descripción general y justificación del curso:
Dada la disponibilidad moderna de grandes cantidades de información que van desde secuencias genómicas y patrones de reactivos químicos hasta información bursátil, se hace necesario el uso y la generación de herramientas que permitan extraer información y dar sentido a esta gran cantidad de datos. Este reto, en el cual la estadística, los modelos matemáticos y la computación hacen parte fundamental del avance en el área del aprendizaje de máquina, tornan fundamental la apropiación de estas herramientas y conceptos.
Objetivo general:

¹ El número de créditos y la intensidad horaria debe estar acorde con el plan de estudios del programa para el que fue diseñado el curso.

Conocer las herramientas y fundamentos estadísticos utilizados en el procesamiento de grandes cantidades de datos y en la implementación de algoritmos de Aprendizaje de Máquina Supervisado (Supervised Machine Learning). Construir modelos estadísticos para predecir o estimar un resultado en base a un conjunto de datos de entrada.

Objetivos específicos:

1. Conocer y aplicar los paquetes de análisis, visualización y manipulación de datos: Pandas, Numpy, Matplotlib.
2. Entender las diferencias entre aprendizaje supervisado y no supervisado, regresión y clasificación.
3. Conocer y comprender los modelos de regresión y aplicarlos a un conjunto de datos.
4. Conocer modelos lineales de clasificación y su uso.
5. Conocer modelos aditivos, de árbol y relacionados.

Contenido:

Unidades:	Temas:	Subtemas:	Horas:
1. Herramientas básicas	Análisis, visualización y manipulación de datos	Uso de los paquetes numpy, matplotlib, pandas	6
2. Introducción al aprendizaje estadístico	¿Qué es el aprendizaje estadístico? Precisión de modelos estadísticos	Mínimos cuadrados y Vecinos Cercanos, Calidad del Fit, Predicción e Interpretabilidad	6
3. Métodos de regresión	Regresión lineal, selección de subconjuntos y reducción	Teorema de Gauss-Markov, Regresión de una y múltiples variables, selección de subconjuntos, regresión de arista y de Lasso.	15
4. Métodos de clasificación	Regresión Logística, Análisis de Discriminante Lineal, K-vecinos cercanos	Modelo logístico, coeficientes de regresión, predicciones, teorema de bayes para clasificación, análisis de discriminante lineal, método de vecinos	15
5. Árboles de decisión	Conceptos básicos, agregación de bootstrap, random forest	Árboles de regresión, árboles de clasificación, comparación de modelos lineales y de árbol, agregación, random forest, boosting.	15
6. Redes Neuronal	Modelo PPR, redes neuronales, redes bayesianas	Modelo Projection Pursuit Regression, ajuste de redes, entrenamiento, overfitting, escalamiento, unidades y capas, evaluación de desempeño	15

3. METODOLOGÍA

Clases magistrales donde se revisan los conceptos fundamentales, sesiones de programación en paralelo y laboratorios en clase. Solución de retos y problemas que utilizan las herramientas y conceptos estudiados.

Las actividades de seguimiento están compuestas por actividades en clase, talleres y al finalizar el curso cada estudiante elaborará un proyecto con lo aprendido en el curso donde debe buscar y solucionar un problema con bases de datos de acceso libre.

Actividad de evaluación	Porcentaje	Fecha
Tareas 1: Unidad 1, 2, 3	24%	29 Enero 2021
Tareas 2: Unidad 3, 4, 5	23%	1 de Marzo 2021
Tareas 3: Unidad 5, 6	23%	1 de Abril 2021
Proyecto final del curso: Artículo, exposición	30%	Pulse para escribir una fecha.w

Actividades de asistencia obligatoria²:

Número de faltas de asistencia máxima permitida : **13 horas (equivalente al 20% de las horas del curso en pregrado, si el curso es para posgrado el límite sería 30%)**

Para el caso de las prácticas académicas defina si la totalidad del curso es de asistencia obligatoria.

Bibliografía:

1. https://github.com/hernansalinas/Curso_aprendizaje_estadistico
2. *An introduction to Statistical Learning: with applications in R*. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. Springer; 1st ed. 2013, Corr. 7th printing 2017 edition.
3. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Trevor Hastie, Robert Tibshirani, Jerome Friedman. Springer 2017.
4. *Probability and Statistics for Engineers and Scientists*. R. Walpole et. al.. Ed. 9. Prentice Hall. 2012.

4. Profesores

Nombres y Apellidos	Dependencia	Formación en pregrado y posgrado	Unidad N°	N° Horas	Fechas
Hernán David Salinas Jiménez	Física	Sí	1-6	72	

5. Aprobación del Consejo de Unidad Académica

² De conformidad con el artículo 30 del Acuerdo Superior 432 de 2014, cuando un estudiante supere el 30% de faltas de asistencia en un curso sin causa justificable legalmente, reprobará por inasistencia y se calificará con una nota de cero, cero (0.0)

Aprobado en Acta número del Haga clic aquí o pulse para escribir una fecha..

	Nombre Completo Secretario del Consejo de la Unidad Académica		Firma		Cargo	