

IDENTIFICACIÓN

NOMBRE ESCUELA	ESCUELA DE CIENCIAS
NOMBRE DEPARTAMENTO	Ciencias Matemáticas
ÁREA DE CONOCIMIENTO	MATEMATICAS, ESTADISTICA Y AFINES
NOMBRE ASIGNATURA EN ESPAÑOL	ESTADÍSTICA MULTIVARIADA AVANZADA
NOMBRE ASIGNATURA EN INGLÉS	ADVANCED MULTIVARIATE STATISTIC
CÓDIGO	CM0868
SEMESTRE DE UBICACIÓN	20212
INTENSIDAD HORARIA SEMANAL	2 horas semanales
INTENSIDAD HORARIA SEMESTRAL	36 horas semestral
CRÉDITOS	
CARACTERÍSTICAS	Suficientable

2. JUSTIFICACIÓN DEL CURSO

El aprendizaje estadístico puede definirse como un conjunto de técnicas y conceptos para entender un conjunto (generalmente grande) de datos. En los últimos años la facilidad de uso y acceso a redes de comunicación ha causado un aumento considerable en la complejidad y la cantidad de información disponible. Tal auge ha provocado que el aprendizaje estadístico se convierta en una herramienta fundamental en la modelación estadística con aplicaciones en muchas y diversas áreas de la ciencia, las finanzas y la industria. Algunas aplicaciones a mencionar son: Identificación de imágenes, diagnóstico médico, predicción de precios, detección de spam, predicción de tráfico, patrones de consumo, entre otras. Lo anterior no habría sido posible sin los avances teóricos en estadística de grandes dimensiones, la disponibilidad de grandes volúmenes de datos y el rol fundamental de la teoría de la computación en la implementación de algoritmos eficientes para los propósitos del aprendizaje estadístico. Este curso se centrará en tales avances teóricos y algorítmicos con un énfasis en su aplicabilidad a problemas de decisión y clasificación.

3. PROPÓSITO U OBJETIVO GENERAL DEL CURSO

3.1. Presentar los elementos principales de la teoría de aprendizaje estadístico con un enfoque práctico, permitiendo al estudiante comprender, plantear, modelar e implementar problemas de decisión y clasificación que se presentan en situaciones reales.

3.2. OBJETIVOS ESPECÍFICOS

- 3.2.1. Utilizar métodos de aprendizaje supervisado para entrenar clasificadores con un alto grado de aciertos predictivos.
- 3.2.2. Utilizar métodos de validación cruzada (cross validation) como medio de verificación de la potencia de predicción del modelo.
- 3.2.3. Plantear y modelar problemas de clasificación no supervisada.
- 3.2.4. Plantear y modelar problemas de clasificación y decisión

4. COMPETENCIAS BÁSICAS QUE EL ALUMNO ESTARÁ EN CONDICIONES DE LOGRAR:

Habilidad para interpretar, entender y modelar problemas de decisión y de clasificación.

Identificar problemas donde es posible aplicar los conceptos de aprendizaje supervisado y no supervisado.

Justificar formalmente la validez de los modelos propuestos.

Usar el lenguaje de programación Python para implementar las soluciones propuestas dentro del marco teórico del aprendizaje automático.

5. DESCRIPCIÓN ANALÍTICA DE CONTENIDOS: TEMAS Y SUBTEMAS

5.1. INTRODUCCIÓN AL APRENDIZAJE SUPERVISADO

- 5.1.1. El problema de clasificación.
- 5.1.2. Métodos simples de predicción: Regresión lineal y vecinos cercanos.
- 5.1.3. Implementación en Python.
- 5.1.4. Error de predicción y validación cruzada.
- 5.1.5. Error de Bayes y función de costo.

5.2. MÉTODOS CLÁSICOS DE CLASIFICACIÓN

- 5.2.1. Hiperplanos separadores óptimos.
- 5.2.2. Métodos subóptimos: Análisis discriminante, regresión logística y perceptron.
- 5.2.3. Métodos de prototipado con k-medias y mezclas gaussianas.

5.2.4. Método de los vecinos cercanos y su aplicación para la identificación de imágenes.

5.2.5. Taller de ejercicios prácticos.

5.3. MÁQUINAS DE SOPORTE VECTORIAL Y REDES NEURONALES

5.3.1. Máquinas de soporte vectorial. Construcción del clasificador de margen maximal y el problema dual.

5.3.2. Funciones de kernel.

5.3.3. Redes neuronales: Formulación, entrenamiento e inconvenientes del modelo.

5.3.4. Taller de ejercicios prácticos.

5.4. APRENDIZAJE NO SUPERVISADO

5.4.1. Métodos geométricos de clustering: k-medias, k-mediodes.

5.4.2. Métodos probabilísticos de clustering: Mezclas gaussianas.

5.4.3. Métodos de descomposición: Componentes principales y componentes independientes.

5.4.4. El algoritmo de page rank.

5.4.5. Completación de matrices y el problema de netflix.

5.4.6. Problemas aplicados

6. ESTRATEGIAS METODOLÓGICAS Y DIDÁCTICAS:

Exposición por parte del profesor de los temas básicos del curso. Plantear, analizar y resolver problemas. Desarrollo de proyectos computacionales en donde se apliquen los conceptos aprendidos.

7. RECURSOS

7.1. Locativos

Aula de Clase

7.2. Tecnológicos

Video Beam. Se necesita contar con equipos de cómputo para cada alumno con el lenguaje de programación Python instalado y conexión a internet

8. CRITERIOS Y POLÍTICAS DE SEGUIMIENTO Y EVALUACIÓN ACADÉMICA

Parcial 1: 30%

Parcial 2: 30%

Talleres prácticos: 40%

9. BIBLIOGRAFIA GENERAL

Libros

Vapnik, Vladimir N. The Nature of Statistical Learning Theory. 1995. Springer.

Hastie, Trevor & Tibshirani, Robert & Friedman, Jerome. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2008. Springer

Casella, George & Berger, Roger L. Statistical Inference. 2002. DUXBURY. Thomson Learning.

Catoni, Oliver & Picard, Jean. Statistical learning theory and stochastic optimization. 2001. Ecole d Eté de Probabilités de Saint-Flour XXXI.

James, Gareth & Witten, Daniela & Hastie, Trevor & Tibshirani, Robert. An Introduction to Statistical Learning with Applications in R. 2013. Springer.

Phuong Vo.T.H and Martin Czygan. Getting Started with Python Data Analysis. 2015.

Grupo ANFI. "Algoritmos De Aprendizaje Automático." Machine Learning. N.p., n.d.

Grupo ANFI. "CURSO I: INTRODUCCIÓN A LA PROGRAMACIÓN CON PYTHON." Curso Introducción a Python. N.p., n.d.

10. NOMBRE DEL PROFESOR COORDINADOR DE MATERIA Y NOMBRE DE PROFESORES DE LA MATERIA QUE PARTICIPARON EN LA ELABORACIÓN.

Coordinador

Ricardo Restrepo

Participante(s)

Freddy H. Marín

11. REQUISITOS DEL PROCESOS DE ASEGURAMIENTO DE LA CALIDAD

Versión número:

1,0

Fecha elaboración:

2017/06/27

Fecha actualización:

2017/06/29

Aprobación:

CARLOS MARIO DE JESUS VELEZ SANCHEZ

