

Unidad Académica Responsable: Departamento de Informática y Ciencias de la Computación

Programa: Magister en Ciencias de la Computación

I.- IDENTIFICACIÓN

| | | |
|---|---------------------|---------------------------|
| Nombre: Machine Learning probabilístico | | |
| Código: | Créditos: 3 | Créditos SCT: 6 |
| Prerrequisitos: Data science 1 | | |
| Modalidad: presencial | Calidad: básica | Duración: semestral |
| Trabajo Académico: 160 | | |
| Horas Teóricas: 32 | Horas Prácticas: 16 | Horas Laboratorio: 16 |
| Horas de otras actividades: 96 | | |
| Horas presenciales: 64 | | Horas no presenciales: 96 |

II.- DESCRIPCIÓN

Esta asignatura proporciona conocimientos avanzados en el área de ciencia de datos y teoría de la probabilidad. Aporta conocimientos avanzados en el análisis estadístico de datos masivos y aprendizaje computacional.

En particular, la asignatura se enfoca en la enseñanza de técnicas avanzadas de modelamiento de datos y muestreo basadas en la teoría de probabilidades tocando, en la parte final, el tema de la causalidad.

Esta asignatura aporta a la siguiente competencia del perfil de egreso:

- Mostrar un manejo profundo y actualizado en Ciencias de la Computación, centrándose en conocimientos fundamentales en teoría de computación.
- Mostrar un conocimiento profundo de la teoría de probabilidades y sus aplicación a la interpretación de resultados científico y al modelamiento de datos.

III.- RESULTADOS DE APRENDIZAJE ESPERADOS

Al finalizar el curso los alumnos deben ser capaces de:

1. Visualizar y explorar grandes volúmenes de datos de diversos tipos.
2. Construir modelos estadísticos entendiendo sus alcances y limitaciones.
3. Aplicar herramientas avanzadas de análisis de datos a problemas abiertos.
4. Desarrollar un proyecto de ciencia de datos utilizando herramientas computacionales avanzadas.

IV.- CONTENIDOS

1. Elementos de Teoría de la Probabilidad
 - a. Probabilidades
 - b. Teorema de Bayes
2. Modelos gráficos
 - a. Redes Bayesianas

- b. Independencia condicional
- c. Campos aleatorios de Markov
- d. Inferencia en modelos gráficos
- 3. Muestreo
 - a. Algoritmos básicos de muestreo: distribuciones estándar, rejection sampling, importance sampling
 - b. Markov Chain Monte Carlo: Metropolis-Hasting, Gibbs
- 4. Inferencia aproximada
 - a. Inferencia variacional
 - b. Regresión lineal variacional
 - c. Distribución de familias exponenciales
 - d. Métodos variacionales locales
 - e. Regresión logística variacional
- 5. Causalidad
 - a. Modelos causales
 - b. Aprendizaje de modelos causales
 - c. búsquedas contrafactuales

V.- METODOLOGÍA

El curso contará con clases teóricas y prácticas. Se requerirá la participación activa de los alumnos mediante la realización de tareas orientadas a la aplicación de la teoría e implementación de distintos algoritmos, el desarrollo de proyectos, la discusión de materiales y la presentación de temas afines.

VI. EVALUACIÓN

La evaluación de la asignatura se calculará como el promedio ponderado de tareas, certámenes, participación en clases, presentación de artículos científicos y un proyecto que se llevará a cabo durante el transcurso del curso.

$$\text{Nota} = 0.1 * \text{Presentación_artículo} + 0.2 * \text{Tarea_1} + 0.2 * \text{Tarea_2} + 0.5 * \text{Proyecto_científico}$$

Si Nota \geq 4.0:

$$\text{Nota_final} = \text{Nota}$$

Si Nota $<$ 4.0:

$$\text{Nota_final} = 0.7 * \text{Nota} + 0.3 * \text{examen_recuperación}$$

Mínimo de notas requeridas: 4. Si el/la estudiante tiene un número de evaluaciones inferior a 4 será considerado/a NCR.

VII.- BIBLIOGRAFÍA Y MATERIAL DE APOYO

Básica

Christopher M. Bishop: Pattern Recognition and Machine Learning. Springer, 2007. ISBN-10: 0387310738, ISBN-13: 978-0387310732.

D. Koller: Probabilistic Graphical Models, Mit Pr (31 July 2009), ISBN-10 : 0262013193

Complementaria

Kevin P. Murphy: Machine Learning: A Probabilistic Perspective. The MIT Press;, 1 edition, 2012. ISBN-10: 0262018020, ISBN-13: 978-0262018029.