

## PRONTUARIO

- A. Título del curso:** Aprendizaje Automático
- B. Codificación del Curso:** CDAT 4011
- C. Cantidad de Horas/Créditos:** Tres (3) horas contacto<sup>1</sup> / Tres (3) créditos
- D. Horas Contacto:** Cuarenta y cinco (45) horas por semestre
- E. Prerrequisitos:** COMP 3082 (Introducción a la Programación y la Ciencia de Cómputos II), ESMA 3105 (Estadística Matemática) o cursos similares previa aprobación de la dirección del Departamento de Matemáticas
- F. Corequisito:** CDAT4013-Laboratorio de Aprendizaje Automático

### G. Descripción del Curso

El aprendizaje automático es la disciplina que busca dotar a las computadoras de la capacidad de aprender a realizar nuevas tareas sin haber sido programadas explícitamente para realizarlas. Esta disciplina opera en la intersección de la matemática, la estadística y la ciencia de cómputos. Su objeto de estudio es la elaboración de algoritmos destinados a construir modelos matemáticos a partir de muestras de datos. En este curso el o la estudiante abordará de forma panorámica fundamentos, técnicas y algoritmos de la disciplina; y se familiarizará de forma práctica, a través de aplicaciones, con el proceso del aprendizaje automático. El contenido del curso se divide principalmente en base a los 3 tipos de aprendizaje considerados en la disciplina: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo.

### H. Objetivos de aprendizaje

Al finalizar el curso el o la estudiante podrá:

1. Distinguir cuál es el enfoque más apropiado de aprendizaje automático para resolver un problema específico.
2. Preprocesar e integrar fuentes heterogéneas de datos, produciendo una entrada de datos expedita para un algoritmo de aprendizaje automático.
3. Producir código eficiente para aplicar los principales algoritmos de aprendizaje automático a problemas específicos.

---

<sup>1</sup> Una hora contacto equivale a cincuenta (50) minutos

4. Evaluar el rendimiento de los modelos producidos bajo diferentes criterios de teoría de aprendizaje.
5. Usar algoritmos más recientes o más especializados de aprendizaje automático.

Al terminar la discusión de cada sección del bosquejo de contenido del curso el o la estudiante podrá:

## I. Introducción

1. Identificar los conceptos básicos necesarios para el estudio de algoritmos de aprendizaje automático
2. Identificar, instalar y configurar las herramientas y entornos de desarrollo básico usados en el aprendizaje automático

## II. Aprendizaje supervisado

1. Describir el problema del aprendizaje supervisado.
2. Describir el problema de regresión.
3. Establecer la diferencia entre una regresión lineal y una logística.
4. Aplicar algoritmos de regresión.
5. Describir el problema de clasificación.
6. Aplicar algoritmos de clasificación.

## III. Teoría de aprendizaje

1. Describir el dilema sesgo / varianza.
2. Explicar el impacto que tiene el dilema sesgo / varianza en la capacidad de generalización de los modelos producidos por los algoritmos de aprendizaje supervisado.
3. Aplicar métodos de validación para mejorar la capacidad de generalización de los modelos.
4. Reducir el sobreajuste a los datos de entrenamiento
5. Aplicar diferentes métricas para resumir.
6. Evaluar el rendimiento de los modelos producidos por la diferentes métricas.

## IV. *Boosting* y *Bagging*

1. Aplicar los meta-algoritmos de *Boosting* y *Bagging* para producir un clasificador robusto a partir de varios clasificadores débiles, tales como árboles de decisión

## V. Redes neuronales

1. Explicar la inspiración biológica de las redes neuronales.
2. Plantear el modelo matemático para representar una neurona.
3. Explicar el algoritmo de entrenamiento del Perceptrón.
4. Describir una red neuronal multicapa.
5. Explicar el algoritmo de entrenamiento de Retropropagación.
6. Entrenar redes neuronales multicapa y ajustar sus hiperparámetros.
7. Identificar otras arquitecturas importantes de redes neuronales de aprendizaje profundo.

## VI. Aprendizaje No Supervisado

1. Describir el problema del aprendizaje no supervisado.
2. Identificar las principales métricas de similaridad.
3. Aplicar algoritmos de conglomerados.
4. Interpretar resultados de algoritmos de conglomerados.

## VII. Reducción de Dimensionalidad

1. Describir el problema de la *maldición* de la dimensionalidad.
2. Aplicar métodos de reducción de dimensionalidad a un conjunto de datos para identificar un subconjunto de variables principales.

## VIII. Aprendizaje por Refuerzo

1. Identificar los elementos de un problema de aprendizaje por refuerzo.
2. Formular un problema de aprendizaje por refuerzo como un proceso de decisión de Markov.
3. Identificar los elementos involucrados en las ecuaciones de optimalidad de Bellman.
4. Describir el concepto de política y función de valor.
5. Aplicar algoritmos para búsqueda de políticas e iteración de políticas.
6. Aplicar algoritmos para iteración de valores y aproximación de la función de valor.
7. Aplicar algoritmos para aprendizaje de diferencias temporales y aprendizaje Q.

## I. Bosquejo del Contenido y Distribución de Tiempo

### I. Introducción (3 horas<sup>2</sup>)

1. Conceptos básicos: cálculo, álgebra lineal y estadística
2. Herramientas necesarias: Python, librerías y entornos de desarrollo

### II. Aprendizaje Supervisado (9 horas)

1. Regresión lineal
2. Regresión logística
3. Modelos lineales para clasificación
4. Análisis discriminante Gaussiano
5. *Naive Bayes*
6. Máquina de vectores de soporte

### III. Teoría de Aprendizaje (6 horas)

1. Dilema sesgo / varianza
2. Regularización y selección de atributos/modelos

---

<sup>2</sup> Una hora contacto equivale a 50 minutos

### 3. Métricas de Evaluación

#### IV. *Boosting* y *Bagging* (3 horas)

1. Árboles de decisión
2. Bosques aleatorios

#### V. Redes Neuronales (9 horas)

1. Algoritmo del Perceptrón
2. Redes Neuronales Multicapa
3. Algoritmo de Retropropagación
4. Entrenamiento y ajuste de hiperparámetros
5. Otras arquitecturas de aprendizaje profundo

#### VI. Aprendizaje No Supervisado (6 horas)

1. k-means clustering
2. Análisis de agrupamiento jerárquico
3. Mezcla de Gaussianas
4. Algoritmo *Expectation Maximization*

#### VII. Reducción de Dimensionalidad (3 horas)

1. Análisis de factores
2. Análisis de componentes principales (PCA)
3. Análisis de componentes independientes (ICA)

#### VIII. Aprendizaje por Refuerzo (3 horas)

1. Procesos de decisión de Markov
2. Búsqueda de políticas e iteración de políticas
3. Iteración de valores y aproximación de la función de valor
4. Diferencia temporal y aprendizaje Q

La suma de las horas sugeridas es de 42. Las 3 horas restantes serán utilizadas para la ejecución evaluaciones y estrategias instruccionales afines con el curso (ver sección H). Los tópicos en este bosquejo de contenido son aptos para ser ordenados de otras maneras a juicio del docente que imparta el curso en el ejercicio de su libertad de cátedra.

## **J. Técnicas Instruccionales**

Con miras a lograr los objetivos del curso el profesor o la profesora seleccionará entre las siguientes técnicas instruccionales: conferencias, videos, demostraciones, trabajos en grupo, estudios independientes, discusión de asignaciones. Además el profesor o profesora podrá implantar y fomentar otras actividades afines al curso.

## **K. Recursos de aprendizaje e instalaciones mínimas disponibles o requeridos**

La Universidad debe proveer un laboratorio para trabajo independiente de los estudiantes, el equipo electrónico que necesita el profesor o la profesora para impartir la clase, el programado apropiado para el curso (Intérpretes y Compiladores de todos los lenguajes que se utilizarán, herramientas para el desarrollo de aplicaciones) y acceso a INTERNET. Se sugiere al estudiante poseer una computadora portátil con capacidad de ejecutar los compiladores, intérpretes u otros programados utilizados en este curso.

## **L. Técnicas de evaluación**

Las evaluaciones consisten en al menos dos exámenes parciales, asignaciones, un proyecto final y la puntuación final del curso de laboratorio CDAT4MLL. La puntuación del curso de laboratorio compondrá un 33% de la nota final. Para el restante 67% se sugiere pesos relativos de 30% para exámenes parciales, 17% para asignaciones, 20% para el proyecto final.

## **M. Acomodo razonable**

La Universidad de Puerto Rico (UPR) de Humacao reconoce el derecho que tienen los estudiantes con impedimentos a una educación post-secundaria inclusiva, equitativa y comparable. Conforme a la política de la Universidad de Puerto Rico hacia los estudiantes con impedimentos, fundamentada en la legislación protectora federal y local vigente en Puerto Rico, todo estudiante cualificado con impedimentos, tiene derecho a la igual participación de aquellos servicios, programas y actividades que están disponibles para la comunidad universitaria en general. Es decir, todo estudiante que tiene un impedimento de naturaleza física, mental o sensorial que lo limita sustancialmente en una o más actividades principales de la vida, podría tener derecho a recibir acomodos o modificaciones razonables en sus áreas de estudios.

Es importante que, de usted requerir algún tipo de acomodo o modificación razonable en este curso, deberá notificar al(a) profesor(a) de su necesidad de acuerdo a la condición física, mental o sensorial que presenta. Además, usted deberá visitar la *Oficina de Servicios para Estudiantes con Impedimentos* (OSEI) del recinto de Humacao para recibir orientación sobre los pasos que deberá seguir para solicitar acomodos o modificaciones razonables y de otros servicios de apoyo que le permitan la obtención de un adecuado aprovechamiento académico. Véase Certificación 133 (2015-2016) de la Junta de Gobierno de la UPR.

## **N. Integridad académica**

La Universidad de Puerto Rico promueve los más altos estándares de integridad académica y científica. El Artículo 6.2 del Reglamento General de Estudiantes de la UPR (Certificación Número. 13, 2009-2010 de

la Junta de Síndicos) establece que “*la deshonestidad académica incluye, pero no se limita a: acciones fraudulentas, la obtención de notas o grados académicos valiéndose de falsas o fraudulentas simulaciones, copiar total o parcialmente la labor académica de otra persona, plagiar total o parcialmente el trabajo de otra persona, copiar total o parcialmente las respuestas de otra persona a las preguntas de un examen, haciendo o consiguiendo que otro tome en su nombre cualquier prueba o examen oral o escrito, así como la ayuda o facilitación para que otra persona incurra en la referida conducta*” . Cualquiera de estas acciones estará sujeta a sanciones disciplinarias en conformidad con el procedimiento disciplinario establecido en el Reglamento General de Estudiantes de la UPR vigente.

#### **O. Normativa sobre discrimen por sexo y género en modalidad de violencia sexual<sup>1</sup>**

“La Universidad de Puerto Rico prohíbe el discrimen por razón de sexo y género en todas sus modalidades, incluyendo el hostigamiento sexual. Según la Política Institucional contra Hostigamiento Sexual, Certificación Núm. 130 (2014-1 5) de la Junta de Gobierno, si un(a) estudiante es o está siendo afectado por conductas relacionadas a hostigamiento sexual, puede acudir a la Oficina de la Procuraduría Estudiantil, el Decanato de Estudiantes o la Coordinadora de Cumplimiento con Título IX para orientación y/o para presentar una queja”.

#### **P. Sistema de calificación**

Se adjudicará la calificación A, B, C, D, ó F según el nivel de competencia demostrado en las evaluaciones. El profesor podrá usar la escala 100-85 A, 84-75 B, 74-60 C, 59-50 D, 49-0 F u otra que resulte más apropiada para las calificaciones del curso y la informará, las primeras semanas de clases (y en la guía de estudiante), la curva a usarse para asignar las calificaciones.

#### **Q. Bibliografía**

1. Hastie, T., R. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, 2016
2. Theobald, O. (2017). *Machine Learning for Absolute Beginners. A Plain English Introduc.*
3. Raschka, S., & Mirjalili, V. (2017). *Python machine learning*. Packt Publishing Ltd.
4. Géron, A. (2017). *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, Inc.
5. Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning*. Cambridge: MIT press
6. Chollet, F. (2017). *Deep learning with Python*. Manning Publications Co.
7. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press
8. Domingos, P. (2015). *The master algorithm: How the quest for the ultimate learning machine will remake our world*. Basic Books

## **R. Créditos<sup>ii</sup>**

Primera versión por Dr. Ollantay Medina Huamán con el insumo de Dr. Elio Ramos Colón, Profa. Idalyn Ríos Díaz y Prof. José O. Sotero Esteva, enero de 2019. Revisión por Comité de Currículo abril 2019, Endosado por Depto. De Matemáticas en RO 12 abril de 2019. Revisión general por Profa. Bárbara L. Santiago-Figueroa, 31 de julio de 2019

Se reconoce la influencia del prontuario *Machine Learning: CS229 Autumn 2018* por Ron Dror y Andrew Ng de Stanford University.

---

### **<sup>i</sup> Traducción del texto:**

“The University of Puerto Rico prohibits discrimination based on sex, sexual orientation, and gender identity in any of its forms, including that of sexual harassment. According to the Institutional Policy Against Sexual Harassment at the University of Puerto Rico, Certification Num. 130, 2014-2015 from the Board of Governors, any student subjected to acts constituting sexual harassment, may turn to the Office of the Student Ombudsperson, the Office of the Dean of Students, and/or the Coordinator of the Office of Compliance with Title IX for an orientation and/or formal complaint”.

<sup>ii</sup> Este prontuario es propiedad de la Universidad de Puerto Rico. Sin embargo, su contenido es una creación académica de su autor. Por consiguiente, en el caso de uso de porciones significativas del mismo se espera reconocimiento del autor tal y como se ha hecho en esta sección.