

PRONTUARIO

A. Título del Curso: Laboratorio de Aprendizaje Automático

B. Codificación del Curso: CDAT 4013

C. Cantidad de Horas/Créditos: Tres (3) horas contacto¹ / Un (1) crédito

D. Horas Contacto: Cuarenta y cinco (45) horas por semestre

E. Prerrequisitos: N/A

F. Correquisitos: CDAT 40ML – Aprendizaje Automático

G. Descripción del Curso

Este curso de laboratorio complementa la formación obtenida en el curso CDAT 4011 – Aprendizaje Automático. Aquí el o la estudiante se familiariza de forma práctica con las diferentes etapas del Aprendizaje Automático y su aplicación a la resolución de problemas reales. Adquiere conocimiento y desarrolla destrezas para la implementación y adaptación de algoritmos de Aprendizaje Automático aplicados a conjuntos de datos específicos. Se familiariza con las técnicas de evaluación de los modelos producidos por los algoritmos de Aprendizaje Automático.

H. Objetivos de aprendizaje

Al finalizar el curso el o la estudiante podrá:

1. Aplicar algoritmos de aprendizaje automático a problemas específicos.
2. Distinguir entre las diferentes técnicas utilizadas en el aprendizaje automático y entender la utilidad particular de dichas técnicas.
3. Ejecutar las diferentes etapas dentro del aprendizaje automático al momento de resolver un problema específico tales como: preprocesamiento de datos, preparación de datos, selección de algoritmo, entrenamiento, ajuste de parámetros y evaluación del modelo.

Al completar cada sección del bosquejo de contenido del curso, el o la estudiante podrá:

1. Reconocer las principales herramientas usadas en el aprendizaje automático.
2. Tener un dominio básico de las principales herramientas usadas en el aprendizaje automático.
3. Explorar un conjunto de datos y prepararlo para aprendizaje supervisado.
4. Entender las diferencias entre un problema de regresión y uno de clasificación.

¹ Una hora contacto equivale a cincuenta (50) minutos

5. Entender las diferencias entre un problema binario y uno multiclase.
6. Producir un clasificador robusto usando un comité de clasificadores.
7. Hacer inferencias sobre datos utilizando aprendizaje no supervisado.
8. Reducir la complejidad de un problema a través de reducción de dimensionalidad de los datos.
9. Estar familiarizado con técnicas de redes neuronales y aprendizaje profundo.
10. Tener un entendimiento básico de la generación de agentes autónomos para la toma de decisiones a través del aprendizaje por refuerzo.

I. Bosquejo del Contenido y Distribución de Tiempo

Cada tema de contenido se aborda en una sesión de tres horas continuas.

I. Instalación de programado (3 horas)

Instalación y configuración de Python y Ambiente de Desarrollo Integrado. Instalación e introducción a las principales librerías para aprendizaje automático, se sugiere *NumPy*, *SciPy*, *scikit-learn*, *matplotlib* y *pandas*. Instalación de *Jupyter Notebook*, herramienta *web* para ejecución de cuadernos computacionales.

II. Regresión lineal (3 horas)

Predicción de una variable continua usando regresión lineal. Implementación de un algoritmo de descenso de gradiente para obtener los parámetros del modelo de regresión.

III. Regresión logística (3 horas)

Predicción de una variable binaria usando regresión logística. Seleccionar un conjunto de datos con dos clases, estandarizarlo y aplicarle un algoritmo de regresión logística.

IV. Clasificación usando clasificadores lineales (3 horas)

Clasificar un conjunto de datos multiclase usando análisis de discriminante lineal.

V. Clasificación usando máquina de vectores de soporte (3 horas)

Clasificar un conjunto de datos multiclase usando máquina de vectores de soporte que incorporen un *kernel* lineal y no lineal.

VI. Clasificación con árboles de decisión y bosques aleatorios (3 horas)

Clasificar un conjunto de datos multiclase usando un árbol de decisión. Visualizar el árbol resultante e interpretarlo. Modificar el proceso de entrenamiento para crear un bosque aleatorio. Comparar las tasas de error entre el árbol de decisión y el bosque aleatorio.

VII. Aprendizaje no supervisado usando conglomerado jerárquico (3 horas)

Realizar un conglomerado jerárquico sobre un conjunto de datos con atributos numéricos reales. Determinar el número de grupos a obtener a través del método del codo. Utilizar al menos dos distancias distintas y comparar resultados.

VIII. Reducción de dimensionalidad (3 horas)

Aplicar análisis de componentes principales a un conjunto de datos para reducir su dimensionalidad. Comparar tasas de error de clasificación después de aplicar un algoritmo de clasificación utilizando el conjunto de datos original y un número reducido de componentes principales.

IX. Instalación de programado para redes neuronales (3 horas)

Instalación, configuración e introducción a un *framework* para aprendizaje profundo, se sugiere *Tensorflow*. Introducción a un *framework* de alto nivel para aprendizaje profundo, se sugiere *Keras*.

X. Redes neuronales (3 horas)

Clasificación de un conjunto de datos utilizando un Perceptrón. Clasificación del mismo conjunto de datos utilizando un red multicapa. Ajuste de hiperparámetros y comparación de resultados.

XI. Otras arquitecturas de redes neuronales (3 horas)

Clasificación de un conjunto de datos con imágenes usando redes de convolución. Clasificación de un conjunto de datos con secuencias usando redes recurrentes

XII. Aprendizaje por refuerzo (3 horas)

Resolver un ambiente virtual, se sugiere OpenAI Gym, usando búsqueda de políticas.

XIII. Periodos para exámenes (2) (6 horas²)

XIV. Presentación de proyecto final de curso (3 horas)

El orden de las experiencias será acoplado al orden en que se presenten los tópicos en el curso CDAT_40ML.

² Se requiere exámenes de tres horas para poder incorporar a la evaluación ejercicios de programación que, por más sencillos que sean, requerirán al menos 30 minutos cada uno para ser completados.

J. Técnicas Instruccionales

Con miras a lograr los objetivos del curso el profesor o la profesora seleccionará entre las siguientes técnicas instruccionales: introducción breve de los temas, demostraciones, trabajo de programación en el laboratorio (individual o en grupo).

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: dos exámenes, trabajos de laboratorio y un proyecto final. Se sugiere pesos relativos de 20% para exámenes, 60% trabajos de laboratorio y 20% para el proyecto final. El puntaje final compondrá un 33% de la nota final del curso CDAT 40ML.

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 sexualⁱ

“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, o F según el nivel de competencia demostrado en las evaluaciones. El profesor o profesora 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. Raschka, S., & Mirjalili, V. (2017). Python machine learning. Packt Publishing Ltd.
3. Géron, A. (2017). Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. O'Reilly Media, Inc..
4. Chollet, F. (2017). Deep learning with Python. Manning Publications Co..

R. Créditosⁱⁱ

Primera versión por Dr. Ollantay Medina Huamán con el insumo del 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

ⁱ 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”.

ⁱⁱ Este prontuario es propiedad de la Universidad de Puerto Rico. Sin embargo, su contenido es una creación académica de su autor. Siguiendo el rigor académico se espera reconocimiento adecuado en caso de uso de porciones significativas tal y como lo han hecho su autor en esta sección.