

Curso: Tópicos Especiales de Ciencia de la Computación

Machine Learning para la Ciencia de Datos

Ciclo: 2019.2

I. INFORMACIÓN GENERAL

HORAS POR SEMANA	: 6 (Teoría: 3, Laboratorio: 3)
CONDICIÓN	: Electivo
PROFESOR	: Juan Espejo
E-MAIL	: jespejod@uni.edu.pe

II. SUMILLA

Actualmente los datos se generan en volúmenes masivos y se vuelve totalmente tedioso para un científico de datos trabajar con ellos. Es entonces cuando el Machine Learning (ML) entra en acción. El ML es la capacidad otorgada a un sistema para aprender y procesar conjuntos de datos de forma autónoma sin intervención humana. El curso introduce al estudiante los fundamentos del ML en vista de su aplicación a la Ciencia de Datos a través de algoritmos básicos y de uso frecuente y técnicas como regresión, agrupación supervisada, clasificadores Bayesianos ingenuos, análisis de componentes principales, entre otras más.

III. COMPETENCIAS

Al finalizar la asignatura, el estudiante:

1. Relaciona el ML con la estadística y el análisis de datos.
2. Utiliza algoritmos informáticos del ML para buscar patrones en los datos.
3. Utiliza técnicas algorítmicas básicas y de uso frecuente que incluyen clasificación, búsqueda, algoritmos voraces y programación dinámica.
4. Entrena algoritmos usando datos de entrenamiento para que pueda predecir el resultado para futuros conjuntos de datos.
5. Conoce, identifica y emplea apropiadamente la teoría matemática (conceptos, teoremas y métodos) que garantiza el funcionamiento de dichos algoritmos.
6. Diseña la inteligencia a través de algoritmos y la computación bioinspirada.

IV. UNIDADES DE APRENDIZAJE

1. Introducción a Python

Obteniendo e instalando Anaconda
Extendiendo Python usando NumPy
Manipulando datos tabulares usando Pandas
Visualización de datos utilizando matplotlib
Referencia: capítulos 1-4 de [2]

2. Comenzando con Scikit-learn para el ML

Obtención de conjuntos de datos
Primeros pasos con Scikit-learn
Limpieza de datos
Referencia: capítulo 5 de [2]

3. Panorama del ML

Tipos de sistemas del ML
Principales desafíos del ML
Prueba y validación
Referencia: capítulo 1 de [1]

4. Bosquejando un Proyecto de ML

Obtención de información de los datos
Preparación de datos para algoritmos de ML
Selección de modelo y entrenamiento
Referencia: capítulo 2 de [1]

5. Aprendizaje Supervisado:

Regresión Lineal

Regresión lineal
Regresión polinomial
Referencia: capítulo 6 de [2] y capítulo 4 de [1]

6. Aprendizaje Supervisado:

Regresión Logística

¿Qué es la regresión logística?
Uso de la base de datos del cáncer de mama de Wisconsin
Referencia: capítulo 7 de [2] y capítulo 4 de [1]

7. Aprendizaje Supervisado:

Clasificación Usando Máquinas de Soporte Vectorial (SVM)

¿Qué son SVM?

Método kernel

Tipos de kernels

Usando SVM para problemas de la vida real

Referencia: capítulo 8 de [2] y capítulo 5 de [1]

8. Aprendizaje Supervisado:

Clasificación Usando k Vecinos más Próximos (k-NN)

¿Qué son k-NN?

Implementación de k-NN en Python

Usando la clase `KNeighborsClassifier` para k-NN

Referencia: capítulo 9 de [2]

9. Árbol de Decisión

Entrenando y visualizando un árbol de decisión

Estimando probabilidades de clase

El algoritmo de entrenamiento CART

Referencia: capítulo 6 de [1]

10. Ensemble Learning y Bosques Aleatorios

Clasificadores de votación

Embolsado y pegado

Parches aleatorios y subespacios aleatorios

Bosques aleatorios

Referencia: capítulo 7 de [1]

11. Reducción de Dimensionalidad

La maldición de la dimensionalidad

Principales enfoques para la reducción de la dimensionalidad

Análisis de Componentes Principales (PCA)

Referencia: capítulo 8 de [1]

12. Aprendizaje No Supervisado:

Agrupamiento Usando k-medias

¿Qué es Aprendizaje No Supervisado?

Usando k-means para resolver problemas de la vida real

Referencia: capítulo 10 de [2] y capítulo 9 de [1]

13. Introducción a las Redes Neuronales Artificiales (ANN) con Keras

De las neuronas biológicas a las artificiales

Implementando perceptrón multicapa (MLP) con Keras

Referencia: capítulo 10 de [1] y capítulo 4 de [3]

14. Desplegando Modelos de ML

Desplegando ML

Caso de estudio

Desplegando el modelo

Referencia: capítulo 12 de [2]

V. CALENDARIO ACADÉMICO

Semana	Fecha	Descripción
1	19/08/19 – 23/08/19	Prueba de entrada / Organización del curso / Unidad 1
2	26/08/19 – 30/08/19	Unidad 2
3	02/09/19 – 06/09/19	Unidad 3
4	09/09/19 – 13/09/19	Unidad 4
5	16/09/19 – 20/09/19	Unidad 5
6	23/09/19 – 27/09/19	Unidad 6
7	30/09/19 – 04/10/19	Unidad 7
8	07/10/19 – 12/10/19	Examen Parcial
9	14/10/19 – 18/10/19	Unidad 8
10	21/10/19 – 25/10/19	Unidad 9
11	28/10/19 – 01/11/19	Unidad 10
12	04/11/19 – 08/11/19	Unidad 11
13	11/11/19 – 15/11/19	Unidad 12
14	18/11/19 – 22/11/19	Unidad 13
15	25/11/19 – 29/11/19	Unidad 14
16	02/12/19 – 07/12/19	Examen Final
17	09/12/19 – 13/12/19	Aniversario de la Facultad de Ciencias
18	16/12/19 – 21/12/19	Examen Sustitutorio

VI. METODOLOGÍA

Se desarrollan sesiones de teoría y laboratorio de cómputo. En las sesiones de teoría, el docente presenta los conceptos, el fundamento de los algoritmos, métodos y técnica. A lo largo de todo el curso, se desarrollará la teoría necesaria de la estadística, el álgebra lineal, la optimización matemática y la teoría de la probabilidad con el fin de dotar al estudiante de un sólido fundamento científico. En las sesiones de laboratorio, se presentará el desarrollo –en el lenguaje de programación Python– de las diversas aplicaciones presentadas en teoría. En todas las sesiones se promueve el aprendizaje colaborativo y el trabajo en equipo.

Referencias

- [1] GÉRON, A. *Hands on Machine Learning with Scikit Learn Keras and TensorFlow*. O'Reilly Media, 2019.
- [2] LEE, W.-M. *Python Machine Learning*. Wiley, 2019.
- [3] SKANSI, S. *Introduction to Deep Learning: From Logical Calculus to Artificial Intelligence*. Undergraduate Topics in Computer Science. Springer, 2018.

Nota importante: interesados deben acercarse a firmar en Escuelas Profesionales, preguntar a la señorita Andrea Hidalgo.