

Escuela de Verano de Aprendizaje Automático

ReMMatEC – PUCE

Septiembre – 2024

1. Datos del evento

- 1. Fechas:** del 9 al 13 de septiembre de 2024
- 2. Horario:** de 8:00 a 10:00 y de 10:30 a 12:30
- 3. Modalidad:** Híbrida
- 4. Lugar:** PUCE campus matriz, Av. 12 de Octubre y Vicente Ramón Roca
 - **Sesión teórica:**
 - **Sesión práctica:**
- 5. Enlace de conexión:**
- 6. Dirigido a:** Estudiantes universitarios de carreras afines al área del Aprendizaje Automático
- 7. Evento gratuito**

2. Cronograma

2.1. Minicursos

Hora	Lunes 9	Martes 10	Miércoles 11	Jueves 12
8:00 – 10:00	Curso 1 sesión teórica	Curso 2 sesión teórica	Curso 3 sesión teórica	Curso 4 sesión teórica
10:30 – 12:30	Curso 1 sesión práctica	Curso 2 sesión práctica	Curso 3 sesión práctica	Curso 4 sesión práctica

2.2. Charlas magistrales

Hora	Viernes 13	Responsable
08:00-08:10	Bienvenida	PUCE
08:10-09:00	Aplicación de Redes Neuronales Informadas en Física (PINNs) en la Reconstrucción de Imágenes Médicas	Evelyn Cueva
09:00-09:50	Charla 2	
10:10-11:00	Charla 3	Golden
11:00-11:50	Charla 4	
11:50-12:00	Cierre del evento	PUCE

3. Cursos

3.1. Curso 1: Fundamentos de Aprendizaje Automático

Temas

Curso Teórico (2 horas)

4. Introducción al Aprendizaje Automático (45 minutos)

- Definición y tipos de aprendizaje automático.
- Aplicaciones y casos de uso.
- Conceptos clave.

2. Preprocesamiento de Datos (30 minutos)

- Importancia del preprocesamiento.
- Métodos de limpieza y transformación de datos.
- Normalización y estandarización.
- División del conjunto de datos.

3. Modelos de Aprendizaje Supervisado (30 minutos)

- Introducción a los modelos de regresión y clasificación.
- Algoritmos supervisados: regresión lineal, regresión logística y árboles de decisión.

4. Evaluación de Modelos (15 minutos)

- Métricas de evaluación.
- Validación cruzada.
- Overfitting y underfitting.

Curso Teórico (2 horas)

1. Introducción a las Herramientas y Configuración (15 minutos)

- Instalación y configuración de Jupyter Notebooks y librerías necesarias (scikit-learn, pandas, numpy, matplotlib).

2. Preprocesamiento de Datos (30 minutos)

- Ejemplos prácticos de limpieza y transformación de datos.
- Normalización y estandarización de datasets.
- División de datasets en conjuntos de entrenamiento y prueba.

3. Implementación de Modelos Supervisados (45 minutos)

- Implementación de una regresión lineal en Python.
- Implementación de una regresión logística en Python.
- Implementación de un árbol de decisión en Python.

4. Evaluación de Modelos (30 minutos)

- Cálculo de métricas de evaluación.
- Implementación de validación cruzada.
- Detección y solución de overfitting y underfitting.

Prerrequisitos

- Conocimientos básicos de Python (variables, tipos de datos, estructuras de control, funciones).
- Familiaridad con el uso de Jupyter Notebooks.
- Conocimientos básicos de álgebra lineal y estadística.

Instructor

- **Kevin Rojas:** Matemático por la USFQ, con una Maestría en Ciencia de Datos y otra en curso en Inteligencia Artificial. Además, cuenta con una Maestría en Economía. Ha sido docente en la USFQ, ESPE, PUCE y UCE, impartiendo clases que combinan teoría y práctica en ciencia de datos. En el sector financiero, ha trabajado en la Junta de Política y Regulación Financiera y en el Banco Diners Club, implementando soluciones basadas en datos. Como consultor, ha asesorado a diversas organizaciones nacionales e internacionales en estadística y análisis de datos. Actualmente, además de la docencia, es CTO de Relatividad IA, donde lidera la implementación de soluciones de inteligencia artificial para empresas. Su enfoque educativo y profesional se centra en traducir conceptos complejos en aplicaciones prácticas, inspirando a sus estudiantes y clientes a aprovechar el poder del análisis de datos y la inteligencia artificial.

4.1. Curso 2: Datos secuenciales

Temas

1. Introducción a los Datos Secuenciales y PyTorch

- Definición y ejemplos de datos secuenciales.
- Importancia y aplicaciones de los datos secuenciales en inteligencia artificial.
- Introducción a PyTorch: instalación y configuración.
- Estructura básica de un programa en PyTorch.

2. Listas y Tuplas en Python

- Creación y manipulación de listas.
- Ejemplos prácticos y ejercicios.

3. Manipulación de Cadenas de Texto

- Operaciones básicas con cadenas de texto.
- Ejemplos prácticos y ejercicios.

4. Introducción a PyTorch para Datos Secuenciales

- Tensores en PyTorch: creación y operaciones básicas.
- Carga y preprocesamiento de datos secuenciales.
- Ejemplos prácticos y ejercicios.

5. Redes Neuronales Recurrentes (RNN) con PyTorch

- Conceptos básicos de RNN.
- Implementación de una RNN simple en PyTorch.
- Entrenamiento y evaluación de la RNN.
- Ejemplos prácticos y ejercicios.

6. Redes Neuronales LSTM con PyTorch

- Conceptos básicos de LSTM.
- Implementación de una LSTM en PyTorch.
- Comparación entre RNN y LSTM.
- Ejemplos prácticos y ejercicios.

7. Procesamiento de Series Temporales con PyTorch

- Introducción a las series temporales.
- Uso de PyTorch para trabajar con series temporales.

Prerrequisitos

- Conocimientos básicos de Python (variables, tipos de datos, estructuras de control, funciones).
- Familiaridad con el uso de Jupyter Notebooks.
- Conocimientos básicos de álgebra lineal y probabilidad.
- Conocimientos básicos de PyTorch (deseable, pero no obligatorio).

Instructor

- **Katherine Morales:** Ingeniera Matemática por la Escuela Politécnica Nacional y master en Ciencia de Datos por École Polytechnique - Francia. Actualmente está culminando su doctorado Télécom SudParis-Francia en el tema: Modelos de Markov generalizados para aprendizaje en datos secuenciales. Su área de estudio son los modelos probabilísticos y técnicas de aprendizaje profundo para aprendizaje supervisado, no supervisado y semi-supervisado para datos secuenciales. Cuenta con experiencia como científica de datos, consultora independiente y capacitadora de la Escuela Politécnica Nacional, la Sociedad Ecuatoriana de Estadística y de la Universidad de los Hemisferios, en temas de aprendizaje supervisado, no supervisado, analítica de datos y manejo de bases de datos.

4.2. Curso 3: NLP

Temas

Parte Teórica

1. Preprocesamiento del Texto (15 minutos)

- Tokenización: palabras, oraciones.
- Normalización del texto: minúsculas, lematización, stemming.
- Eliminación de ruido: stop words, puntuación, caracteres especiales.

2. Representación del Texto (15 minutos)

- Bolsas de palabras (Bag of Words, BoW).
- TF-IDF (Term Frequency-Inverse Document Frequency).
- Word embeddings: Word2Vec, GloVe, FastText.
- Modelos basados en transformers: introducción a BERT y GPT.

3. Modelos de Lenguaje (40 minutos)

- N-gramas.
- Modelos probabilísticos.
- Modelos basados en deep learning: RNN, LSTM, GRU.
- Transformers y su arquitectura básica.

Modelos de Lenguaje de Gran Escala (LLMs) (35 minutos)

- Introducción a LLMs (GPT, BERT, GPT-3, GPT-4).
- Arquitectura y entrenamiento de LLMs.
- Aplicaciones y limitaciones de los LLMs.

Tareas de NLP (15 minutos)

- Clasificación de texto.
- Análisis de sentimientos.
- Traducción automática.
- Generación de texto.

- Resumen automático.
- Reconocimiento de entidades nombradas (NER).

Parte Práctica

1. Preprocesamiento del Texto con Python (15 minutos)

- Importación de librerías: NLTK, SpaCy.
- Tokenización y normalización del texto.
- Eliminación de ruido.
- Ejercicios prácticos.

2. Representación del Texto (15 minutos)

- Implementación de Bag of Words y TF-IDF con `scikit-learn`.
- Uso de embeddings preentrenados con `gensim` (Word2Vec, GloVe).
- Ejercicios prácticos.

3. Construcción de Modelos de Clasificación de Texto (20 minutos)

- Construcción de un clasificador básico usando TF-IDF y un algoritmo de clasificación (Naive Bayes, SVM) con `scikit-learn`.
- Entrenamiento y evaluación del modelo.
- Ejercicios prácticos.

4. Implementación de Modelos Avanzados (30 minutos)

- Uso de modelos preentrenados como BERT con `transformers` de Hugging Face.
- Finetuning de un modelo BERT para una tarea específica (por ejemplo, análisis de sentimientos).
- Evaluación y análisis de resultados.
- Ejercicios prácticos.

5. Introducción a LLM

- Uso de modelos preentrenados como GPT-3 o GPT-4 con `transformers` de Hugging Face.
- Generación de texto con LLMs.

Prerrequisitos

- Conocimientos básicos de Python y machine learning.

Instructor

- **Ana Julia Escobar:** reseña

4.3. Curso 4: Procesamiento de Imágenes

Temas

Parte Teórica (2 horas):

1. Introducción al Procesamiento de Imágenes

- Definición y aplicaciones del procesamiento de imágenes.
- Librerías y herramientas básicas: OpenCV, PIL.
- Fundamentos de la manipulación de imágenes: lectura, escritura y visualización.

2. Preprocesamiento de Imágenes

- Conversión de color y espacios de color.
- Transformaciones geométricas: escalado, rotación, traslación.
- Operaciones aritméticas y lógicas en imágenes.

3. Filtrado y Mejoramiento de Imágenes

- Filtros espaciales: suavizado, detección de bordes.
- Filtros frecuenciales: Transformada de Fourier.
- Técnicas de mejoramiento de contraste: histograma y ecualización.

4. Segmentación de Imágenes

- Técnicas de segmentación: umbralización, detección de contornos.
- Segmentación basada en regiones y bordes.

Parte Práctica (2 horas):

- 1. Manipulación de Imágenes con OpenCV y PIL**
 - Lectura, escritura y visualización de imágenes.
- 2. Transformaciones Geométricas y Operaciones Básicas**
 - Aplicar escalado, rotación y traslación a imágenes.
- 3. Filtrado de Imágenes**
 - Aplicación de filtros espaciales: suavizado, detección de bordes.
- 4. Implementación de una Red Neuronal Convolutiva (CNN) Simple**
 - Creación y entrenamiento de una CNN para la detección de objetos básicos.
 - Evaluación y análisis de resultados.

Prerrequisitos

- Conocimientos básicos de Python.
- Familiaridad con conceptos básicos de álgebra lineal y cálculo.

Instructor

- **Andrés Merino:** Profesor Agregado I en la Pontificia Universidad Católica del Ecuador y Coordinador de la Escuela de Ciencias Físicas y Matemática. Su formación académica integra el título de Matemático de la Escuela Politécnica Nacional con distinción Summa Cum Laude, y es ampliada a través de una Maestría en Matemáticas Puras y Aplicadas por la Universidad Central del Ecuador y un Máster Universitario de Ciencia de Datos por la Universitat Oberta de Catalunya. Ha dedicado su carrera a la docencia y la investigación. En este último, colabora en las áreas de la Teoría Descriptiva de Conjuntos, Fundamentos de la Matemática, Educación Matemática y Ciencia de Datos. Ha aportado conocimientos y avances mediante publicaciones en revistas científicas y participación en congresos nacionales e internacionales. Sus contribuciones incluyen estudios sobre la Derivada de Cantor-Bendixson, Modelos de aprendizaje

automático para análisis de datos e imágenes y Técnicas innovadoras para la enseñanza de la Matemática.

5. Auspiciantes publicitarios

- Sociedad Ecuatoriana de Estadística
- Women in Data Science – Ecuador

6. Lanzamiento

Se espera realizar el lanzamiento a inicios de julio.

