SEMINARIO
REMOTO

Universidad del Rosario | Educación Continua y Consultoría

FUNDAMENTOS EN APRENDIZAJE AUTOMÁTICO CON SCKIT - LEARN

CONTENIDO TEMÁTICO

Módulo 1.

INTRODUCCIÓN Y PASOS EN UN PROYECTO DE APRENDIZAJE AUTOMÁTICO

Definición, propósito y problemas que el aprendizaje automático busca resolver. Categorías principales y conceptos fundamentales de sus sistemas. Descripción general de los pasos típicos en un proyecto de aprendizaje automático.

Módulo 2.

CLASIFICACIÓN

Abordaremos las tareas de aprendizaje supervisado, centrándonos en la clasificación (predicción de clases). Se realizará una revisión de los sistemas de clasificación mediante el análisis de ejemplos con scikit-Learn.

Módulo 3.

REGRESIÓN LINEAL, REGRESIÓN MÚLTIPLE

En este módulo, abordamos los modelos de aprendizaje automático y sus algoritmos de entrenamiento, priorizando la regresión lineal y múltiple. Aunque muchas veces se utilizan como cajas negras, entender su funcionamiento interno permite seleccionar el modelo adecuado, el algoritmo de entrenamiento correcto y un conjunto óptimo de hiperparámetros. Se exploran dos métodos para entrenar la regresión lineal: mediante una ecuación de "forma cerrada" y mediante el enfoque de optimización iterativa llamado descenso del gradiente. Posteriormente, se introduce la regresión polinómica, que aborda conjuntos de datos no lineales y técnicas de regularización para evitar el sobreajuste. El módulo concluye con una breve introducción a la regresión logística y softmax, todo ello implementado con scikit-Learn.

Módulo 4.

SVM

Support Vector Machine(SVM, por sus siglas en inglés) es un modelo versátil y potente de aprendizaje automático, apto para realizar clasificación lineal o no lineal, regresión e incluso detección de novedades. Las SVMs son especialmente efectivas en conjuntos de datos pequeños a medianos de carácter no lineal, principalmente para tareas de clasificación. Sin embargo, no se adaptan eficientemente a conjuntos de datos muy grandes. En este módulo, abordaremos los conceptos esenciales de las SVMs, su uso y su funcionamiento, todo ello a través de ejemplos prácticos utilizando scikit-learn.

Módulo 5.

ARBOLES DE DECISIÓN

Los árboles de decisión son algoritmos versátiles de aprendizaje automático que pueden realizar tareas de clasificación y regresión, e incluso tareas de múltiples salidas. Son algoritmos poderosos, capaces de adaptarse a conjuntos de datos complejos. En este módulo, comenzaremos discutiendo cómo entrenar, visualizar y realizar predicciones con árboles de decisión utilizando scikit-learn. Luego, abordaremos el algoritmo de entrenamiento CART, exploraremos cómo regularizar árboles y cómo utilizarlos para tareas de regresión. Finalmente, discutiremos algunas de las limitaciones de los árboles de decisión.

Módulo 6.

ENSEMBLE LEARNING Y RANDOM FOREST

El "ensemble Learning" aprovecha la sabiduría colectiva, combinando múltiples predictores para mejorar la precisión en tareas de aprendizaje automático. Es similar a consultar a miles de personas para obtener una respuesta más acertada. En este tema, exploraremos técnicas de ensemble, como "random forests", clasificadores de votación y "boosting", y cómo implementarlas con scikit-learn. Estos métodos, al agrupar diferentes algoritmos, potencian las predicciones y son esenciales en proyectos avanzados.

Módulo 7.

REDUCCIÓN DE LA DIMENSIONALIDAD

El enfrentar problemas con miles o millones de características puede llevar a la "maldición de la dimensionalidad", complicando y ralentizando el proceso de aprendizaje automático. Afortunadamente, es posible reducir estas características, facilitando y agilizando la tarea.
La reducción de dimensionalidad no solo acelera el entrenamiento, sino que es crucial para la visualización de datos, facilitando la identificación de patrones y la comunicación efectiva de los resultados. En este módulo, exploraremos técnicas de reducción de dimensionalidad, como PCA, proyección aleatoria y LLE, y cómo implementarlas con scikit-learn para optimizar el aprendizaje automático.

Módulo 8.

TÉCNICAS DE APRENDIZAJE AUTOMÁTICO NO SUPERVISADO

El aprendizaje no supervisado brinda la posibilidad de trabajar con datos no etiquetados, una realidad en muchas aplicaciones actuales de aprendizaje automático. En lugar de depender de etiquetas manuales costosas y laboriosas, este enfoque se centra en tareas como el agrupamiento de instancias similares, la detección de datos inusuales y la estimación de la densidad de un conjunto de datos.
En este módulo, utilizando scikit-learn, abordaremos estos conceptos esenciales, proporcionando herramientas prácticas para situaciones del mundo real.