Apuntes Semana 3 - 04/03/2025

Elaborado por: Raul Sanabria Marroquin - 2020182835

Abstract—Este documento presenta un resumen estructurado sobre Machine Learning, sus conceptos fundamentales, tipos de aprendizaje, el pipeline de desarrollo y métricas de evaluación.

1. Introducción

Machine Learning (ML) es un campo que tiene 2 aristas de importancia: ciencia e ingeniería, algunos ejemplos de usos importantes del machine learning son: construir algoritmos capaces de realizar tareas sin ser explícitamente programados, basándose en inferencias de los datos o crear algoritmos para aproximar funciones, cabe destacar que estas aproximaciones son utiles y no necesariamente perfectas.

2. Machine Learning en Ciencia

ML permite la generación de nuevo conocimiento y desarrollo de investigaciones, contribuyendo a áreas como Data Science y Research Science brindando puestos en estas áreas.

3. Machine Learning en Ingeniería

ML en ingeniería implica la puesta en producción de modelos utilizando herramientas como MLOps y transformaciones de modelo con ONNX. Cabe destacar otros conceptos como la existencia de la inteligencia articial con los algoritmos generativos, el machine learning con los metodos estadisticos y el deep learning con los Large language models.

4. Tipos de Aprendizaje

- **Supervised Learning**: tiene una etiqueta o un data set que da un refuerzo al algoritmo para ver si esta haciendo bien o no las cosas.
- **Unsupervised Learning**: no tiene etiqueta y por ende no hay refuerzo, aparece el clustering.
- **Semi-supervised Learning**: de todo el data set, una parte esta supervisada y la otra no.
- **Auto supervised Learning**: el mismo input actúa como etiqueta.
- **Reinforcement Learning**: se dan recompensas al algoritmo por decisiones correctas.
- **Few-shot Learning**: se le da unos cuantos ejemplos para realizar la tarea.
- **One-shot Learning**: se enseña con un solo ejemplo.
- **Zero-shot Learning**: se le indica al modelo que haga una tarea sin haber sido entrenado en esa área en especifico.

5. Pipeline de Machine Learning

5.1. Data adquisition

¿De donde sacamos los datos? Los datos deben ser relevantes al problema a resolver. La calidad de los datos importa. Ej: Elecciones presidenciales sin tomar en cuenta la región , eso esta MAL!!!

5.2. Data preparation

¿Como le voy a presentar los datos a mi algoritmo? Se eliminan valores duplicados o faltantes, y se normalizan los datos para mantener una distribución adecuada.

5.3. Feature Engineering

Creación y selección de características relevantes para mejorar el modelo a partir de caracteristicas ya existentes, se hace una seleccion de caracteristicas relevantes.

5.4. Model selection

Se elige el mejor modelo según los requisitos del problema.

5.5. Model training

Incluye la optimización de hiperparámetros mediante métodos como GridSearch.

5.6. Model deployment

Proceso de llevar el modelo entrenado a producción.

5.7. Supervised Learning:

Aprendizaje con un conjunto de datos etiquetado. Pares de entrada y salida conocidos. Los modelos son supervisados por los datos. Los datos pueden ser número, texto, imágenes, o audio.

6. Conceptos Clave en Machine Learning

6.1. Feature Vector

Vector en un espacio N-dimensional que representa características de una instancia.

6.2. Label

Información que el modelo debe predecir o clasificar.

6.3. Dataset

Colección de instancias para entrenamiento y evaluación.

6.4. Hiperparámetros

Definidos por el desarrollador, afectan el desempeño del modelo y se ajustan empíricamente.

7. Vectores y Distancias

En el contexto de Machine Learning, los vectores son importantes para representar datos y realizar cálculos distancia entre puntos en un espacio multidimensional.

7.1. Vectores Bidimensionales

Un vector bidimensional es un vector en el que cada instancia tiene dos componentes, representando puntos en un plano cartesiano. Es uno de los casos más simples y frecuentemente utilizado en ejemplos de Machine Learning.

7.2. Distancia L1 - Manhattan

La distancia L1, también conocida como distancia Manhattan, calcula la suma de las diferencias absolutas entre las coordenadas de dos puntos en un espacio N-dimensional. Esta distancia es útil cuando los datos se distribuyen de manera no uniforme.

7.3. Distancia L2 - Euclideana

La distancia L2, también conocida como distancia Euclideana, mide la longitud del segmento recto entre dos puntos en el espacio. Es la distancia "normal" o la más comúnmente utilizada, especialmente en geometría y machine learning.

Fecha de entrega: martes 11 de marzo.