

TRABAJO 1 (Pesa el 15% de los trabajos) - ENTREGA MÁXIMA 2 DE MARZO 23:59

Con el dataset colectado en clase, que contiene las siguientes variables:

- **Altura** (en centímetros)
- **Peso** (en kilogramos)
- **Género**
- **Tamaño de pie**

Utilizarás este dataset para implementar y comparar distintos algoritmos de regresión y clasificación, haciendo uso de **funciones reutilizables** para el cálculo de gradientes, actualización de parámetros, cómputo de predicciones, etc.

Parte 1: Regresión sobre variables numéricas

Objetivo:

Predecir la **altura en centímetros** de la persona utilizando las demás variables (peso, género, tamaño de pie) mediante dos enfoques:

1. **Regresión lineal mediante la solución analítica de mínimos cuadrados.**
 2. **Regresión lineal mediante batch gradient descent.**
-

Parte 2: Clasificación con variable categórica

Objetivo:

Utilizando el mismo dataset, redefine el problema: el **target** será una variable binaria que indique si la altura es **mayor a 1.75 m** o **menor o igual a 1.75 m**.

Implementa los siguientes modelos de regresión logística:

1. **Regresión logística mediante batch gradient descent.**
 2. **Regresión logística mediante stochastic gradient descent.**
-

Entrega

Debes entregar un **notebook** que contenga:

1. **Funciones reutilizables:**
Una o varias celdas con las funciones necesarias (cálculo de gradientes, actualización de parámetros, cómputo de predicciones, etc.).
2. **Escenarios de entrenamiento:**

- Una celda para la implementación y entrenamiento de la **regresión lineal** mediante la solución analítica de mínimos cuadrados.
- Una celda para la implementación y entrenamiento de la **regresión lineal** mediante batch gradient descent.
- Una celda para la implementación y entrenamiento de la **regresión logística** mediante batch gradient descent.
- Una celda para la implementación y entrenamiento de la **regresión logística** mediante stochastic gradient descent.

3. Resultados adicionales:

- Para los dos escenarios de la regresión lineal hacer un assertion de que los parámetros obtenidos son iguales (o casi iguales con un pequeño margen de redondeo)
- Para los dos escenarios de la regresión lineal medir el SSR y el R²
- Para los dos escenarios de la regresión logística hacer un assertion de que los parámetros obtenidos son iguales (o casi iguales con un pequeño margen de redondeo)
- Para los dos escenarios de la regresión logística medir el % de aciertos y la cross entropy loss

4. Output visible:

En cada celda se debe mostrar el output relevante de cada operación:

- Los parámetros obtenidos durante el entrenamiento.
- El resultado del **assert** de validación.
- Las métricas correspondientes: SSR y R² para la regresión; porcentaje de aciertos y cross entropy loss para la clasificación.

5. Conversión a HTML:

El notebook debe estar convertido a formato HTML para la entrega final.

Notas adicionales:

Asegúrate de que el código esté bien comentado y estructurado para facilitar su comprensión y reutilización. Sólo podéis utilizar las librerías Numpy o Pandas, aunque también se puede hacer con objetos primitivos de Python como listas.

¡Buena suerte y manos a la obra!

