**Universidad Mayor Real y Pontificia San Francisco Xavier de Chuquisaca**

**“Facultad de tecnología”**



**Nombre:** Claudia Pereira Cuba

**Carrera:** Ing. Ciencias de la Computación

**Materia:** SIS420

**Laboratorio:** 3

Sucre – Bolivia

Link de github: <https://github.com/clpereirac/SIS420-IA/tree/main/Laboratorios/Lab-3>

**“Regresión logística binaria”**

Primeramente se eligió un dataset que contiene las siguientes características:

|  |  |  |
| --- | --- | --- |
| Nro | Características | Descripción de la característica |
| 1 | PAQ605 | Si el encuestado participa en deportes, actividades físicas o actividades recreativas de intensidad moderada o vigorosa en la semana típica |
| 2 | BMXBMI | Índice de masa corporal del encuestado |
| 3 | LBXGLU | Glucemia del encuestado después del ayuno |
| 4 | LBXGLT | Opinión oral del demandado |
| 5 | LBXIN | Niveles de insulina en sangre del encuestado |
| 6 | Age Group (Target) | ¿El encuestado es una persona mayor o no? |

1. Preprocesamiento

La clase de destino contiene dos valores: 0 o 1, donde 0 se refiere a un adulto y 1 a una persona mayor.

Cuando mencionamos adulto en nuestro contexto actual podemos hacer referencia a una persona comprendida entre lo 18 y 60 años y una persona mayor desde los 60 hasta los 100 años.

En el preprocesamiento del dataset se:

* Comprobó valores faltantes
* Se verificamos si los datos están normalizados (revisamos la: Media y la Desviación Estándar)
* Se verificamos que los datos estén normalizados
* Se busco valores atípicos
* Se elimino valores atípicos
* Se normalizo los datos

Se dividió los datos de entrenamiento y prueba (80% de los datos para entrenamiento y el 20% de prueba)

Implementación del modelo:

Se implementó un modelo de regresión logística utilizando la función sigmoide.

Se utilizó el descenso de gradiente para optimizar los parámetros (theta) del modelo.

Entrenamiento del modelo:

Se aplicó el descenso de gradiente con una tasa de aprendizaje (alpha) de 0.01 y 10,000 iteraciones.

Se obtuvo un conjunto de parámetros theta optimizados.

1. Modelo

Se calculó el costo.

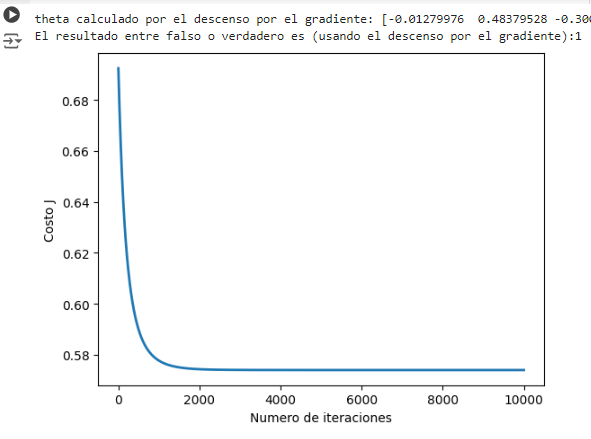
Resultados:

El costo inicial con theta en ceros fue de 0.693.

El costo final después de la optimización fue de 0.584, lo que indica una mejora en el ajuste del modelo.

Se obtuvo una precisión de entrenamiento,

Interpretación del gráfico de costo:



Se observa una curva descendente que muestra cómo el costo disminuye con el número de iteraciones.

Predicciones:

El modelo es capaz de hacer predicciones binarias (adulto o persona mayor) basadas en las características de entrada.

En resumen, se ha implementado y entrenado con un modelo de regresión logística, para predecir si una persona es un adulto (de 18 a 60 años aproximadamente) o una persona mayor (de 60 hasta 100 años) basándonos en características de salud y actividad física. La disminución del costo y la capacidad de hacer predicciones indican que el modelo ha aprendido patrones útiles de los datos.