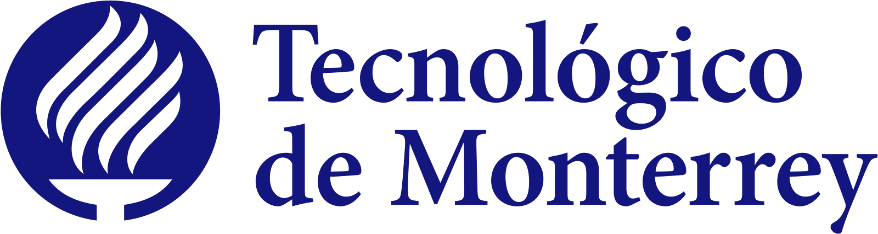
Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Estado de México



Inteligencia artificial avanzada para la ciencia de datos I (Gpo 101)

TC3006C.101

Profesores:

Jorge Adolfo Ramírez Uresti

**Momento de Retroalimentación: Módulo 2 Implementación de una técnica de aprendizaje máquina sin el uso de un framework**

Abner Maximiliano Lecona Nieves | A01753179

01 de Septiembre 2024

Documentación de la Regresión Logística

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

pandas: Para manejar los datos tabulares en los archivos CSV.

numpy: Para operaciones matemáticas y manipulación de matrices.

matplotlib.pyplot: Para graficar la evolución del costo durante el entrenamiento.

Texto

Descripción generada automáticamente

Implementa la función sigmoide que convierte un valor real en una probabilidad entre 0 y 1.

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

Calcula la hipótesis (o predicción) usando los parámetros del modelo y una muestra, aplicando la función sigmoide.

Texto

Descripción generada automáticamente

Normaliza las características de las muestras usando la técnica de Min-Max Scaling, excluyendo el término de sesgo (primera columna).

Texto

Descripción generada automáticamente

Normaliza nuevos datos usando los valores mínimos y rangos calculados previamente.

Texto

Descripción generada automáticamente

Predice si una muestra pertenece a la clase 1 (pasa el examen) o 0 (no pasa), normalizando las características y aplicando la hipótesis.

Texto

Descripción generada automáticamente

Actualiza los parámetros del modelo mediante el descenso de gradiente, ajustando los pesos con base en el error de cada muestra.

Texto

Descripción generada automáticamente

Calcula la función de costo logístico para evaluar el error del modelo en las predicciones.

Texto

Descripción generada automáticamente

Entrena el modelo utilizando descenso de gradiente hasta que los parámetros convergen o se alcance el número máximo de épocas. Se grafica la evolución del costo durante el entrenamiento.

Imagen que contiene Texto

Descripción generada automáticamente

Función para calcular las métricas necesarias para la validación de datos y de información.

Texto

Descripción generada automáticamente

Evalúa el rendimiento del modelo calculando la precisión, recall, F1 Score y la matriz de confusión.

Texto

Descripción generada automáticamente

Se cargan los datos de entrenamiento, validación y prueba.

Se extraen las características y etiquetas.

Se entrena el modelo con los datos de entrenamiento.

Se evalúa el modelo con los datos de prueba y validación.

Interfaz de usuario gráfica

Descripción generada automáticamente con confianza media

Grafica de costo vs épocas (después de 50000 épocas)

Texto

Descripción generada automáticamente

Datos del modelo con el data set de validacion

Texto, Calendario

Descripción generada automáticamente

Datos del modelo con el data set del testing.

Texto

Descripción generada automáticamente

Datos reales del testing.

Ya con el modelo entrenado, validado y probado. Podemos decir que el modelo tiene un excelente desempeñó ya que la precisión de este esta al 97%, Ya con el modelo entrenado, validado y probado, podemos concluir que el modelo tiene un desempeño muy bueno, con una precisión de 94.3%, un recall perfecto del 100% y un F1 Score de 97.1%. Esto indica que el modelo es muy eficaz al identificar correctamente a los estudiantes que pasarán el examen sin perder ningún caso positivo, aunque tiene una pequeña tasa de falsos positivos.

La precisión alta y el recall perfecto sugieren que el modelo no solo predice con exactitud, sino que también es exhaustivo en la identificación de los casos positivos. La matriz de confusión muestra que, aunque existen pocos errores (2 falsos positivos), estos son mínimos comparados con el total de predicciones.