

Instituto Tecnológico y de Estudios Superiores de Monterrey
Inteligencia artificial avanzada para la ciencia de datos I (Gpo 101)

Módulo 2

Implementación de una técnica de aprendizaje máquina sin el uso de un framework.

Alumna:

Mariluz Daniela Sánchez Morales

Profesor:

Jorge Adolfo Ramírez Uresti

1 de Septiembre de 2024

First Run:

```
Final error: 0.09595177992369223
Convergence reached.
Final parameters: [ 7.60179163 62.85344395]

Validation Precision: 1.0
Validation Recall: 1.0
Validation F1 Score: 1.0
Validation Confusion Matrix: (2, 1, 0, 0)

Test Precision: 1.0
Test Recall: 1.0
Test F1 Score: 1.0
Test Confusion Matrix: (3, 2, 0, 0)

Test Set Predictions:
Sample: 1.0, Actual: 0, Predicted: 0
Sample: 1.0, Actual: 1, Predicted: 1
Sample: 1.0, Actual: 1, Predicted: 1
Sample: 1.0, Actual: 0, Predicted: 0
Sample: 1.0, Actual: 1, Predicted: 1
```

Second Run:

```
Final error: 0.1429318825273971
Convergence reached.
Final parameters: [ 7.7484882 63.70966184]

Validation Precision: 1.0
Validation Recall: 1.0
Validation F1 Score: 1.0
Validation Confusion Matrix: (2, 1, 0, 0)

Test Precision: 1.0
Test Recall: 1.0
Test F1 Score: 1.0
Test Confusion Matrix: (2, 3, 0, 0)

Test Set Predictions:
Sample: 1.0, Actual: 1, Predicted: 1
Sample: 1.0, Actual: 0, Predicted: 0
Sample: 1.0, Actual: 1, Predicted: 1
Sample: 1.0, Actual: 0, Predicted: 0
Sample: 1.0, Actual: 0, Predicted: 0
```

Conclusión:

El modelo de regresión logística ha demostrado un buen rendimiento tanto en el set de validación como en el set de prueba, ya que en el primer intento salió con un error final de aproximadamente **0.096**. El modelo logró minimizar la pérdida logística a menos de 0.1, lo que indica un buen ajuste a los datos de entrenamiento.

Las métricas de evaluación de los sets de datos todos resultaron en 1.0, lo que resalta la capacidad del modelo para clasificar correctamente todas las instancias sin falsos positivos ni falsos negativos. Además de que los resultados de las matrices de confusión ayudan a validar el rendimiento de este modelo y sugiere que es confiable en cuanto a las predicciones.

Sin embargo, en el segundo intento, aunque las métricas de evaluación de los sets siguen con los mismos valores sin falsos positivos ni falsos negativos, podemos observar un aumento en el error final con los mismo datos de prueba de *0.096* a **0.143**. Esto se puede atribuir a la repartición de datos en los diferentes sets ya que se maneja de manera aleatoria y esto puede influir negativamente a la precisión del modelo.

En resumen, el modelo de regresión logística sigue mostrando un rendimiento robusto y capacidades de generalización. El ligero aumento en el error debe ser monitoreado, pero no impacta significativamente la efectividad general del modelo. Esto indica que el modelo es confiable para predecir resultados binarios en el conjunto de datos dado. Experimentos adicionales con diferentes divisiones de datos o técnicas de regularización adicionales podrían proporcionar más información sobre su rendimiento y robustez.