

Tarea 03: Implementación de Regresión Logística - PyTorch

Curso de Inteligencia Artificial
Escuela de Ingeniería en Computación
Instituto Tecnológico de Costa Rica

I. OBJETIVO

El propósito de esta tarea es implementar y analizar un modelo de regresión logística para resolver un problema de clasificación binario para determinar si la calidad de un vino es "BUENA" o "MALA", utilizando como entorno de desarrollo PyTorch para la optimización y creación de modelos personalizados.

Se espera que el estudiante realice un análisis exploratorio detallado de las características del conjunto de datos y la variable a predecir, que implemente stratified sampling, ajuste un modelo de regresión logística y evalúe aplicando métricas.

II. PASOS A SEGUIR

A. Descarga del Conjunto de Datos

Descargar el dataset desde el siguiente enlace: Kaggle

B. Análisis de Características

Realizar un análisis exploratorio de datos (EDA) que incluya:

- Realizar un análisis exploratorio de datos (EDA): histogramas y boxplots de variables numéricas, diagramas de dispersión entre características y el target, matriz de correlación (con mapa de calor), o cualquier visualización que considere necesaria.
- Solamente puede seleccionar como máximo 6 características, por lo tanto debe hallar la forma óptima para proceder con el descarte de las sobrantes. Justifique matemáticamente.
- Puede aplicar feature engineering en caso de que se requiera. Especialmente en la variable a predecir.
- Obligatorio dividir el conjunto en: entrenamiento, validación y testing. Seleccione la técnica que mejor se adecue al problema.

El análisis debe incluir no solo visualizaciones, sino también interpretación de los resultados.

C. Implementación de Regresión Logística

- Implementar la regresión logística utilizando PyTorch.
- Una vez definido el modelo ejecute 10 entrenamientos cambiando valores de hiperparámetros (learning rate, batch, epochs)

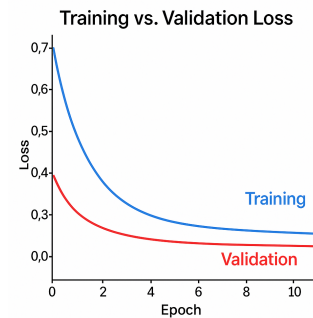


Fig. 1. Training vs Loss sample

D. Evaluación del Modelo

- Debe mostrar los resultados de entrenamiento y validación de acuerdo a la función de pérdida que usted considera conveniente para este problema, como se muestra en la Figura 1. De acá debe determinar si existe overfitting de su modelo en cada entrenamiento.
- Además capture métricas para evaluar el accuracy, precision, recall, F1-Score y AUC, y establezca un cuadro comparativo entre los 10 diferentes entrenamientos.

E. Análisis de Resultados

- Compare el resultado final del entrenamiento con el conjunto de test y muestre los resultados finales de su modelo. Seleccione aquel modelo que dio mejores resultados durante el proceso de entrenamiento
- Genere una matriz de confusión con el set de testing para evaluar los resultados de su modelo.
- Capture nuevamente las métricas de su modelo final seleccionado con el conjunto de testing y discútalas en su informe.

F. Entrega

El informe deberá realizarse en \LaTeX (Overleaf) utilizando la plantilla IEEE para artículos científicos. El documento debe contener visualizaciones, y la interpretación de los resultados. Además, se debe adjuntar un **Jupyter Notebook** con el código implementado. La entrega final consistirá en un archivo comprimido (.zip) que contenga:

- Código fuente en \LaTeX .
- El PDF del informe.
- El Notebook con el código fuente.

RÚBRICA

Criterio	Puntaje Máx.
Desarrollo de Regresión Logística con PyTorch	20
Justificación de selección de features	20
Presentación de métricas correctas por experimento	25
Análisis, comparación de resultados, selección de modelo final	35
Total	100

Si el trabajo no se encuentra debidamente ordenado y presentado siguiendo una adecuada estructura para el informe, puede ser considerado como incompleto y cualquiera de las rúbricas se puede ver afectada.