

Instrucciones: Optimización de LightGBM para Clasificación Binaria

Resumen

Este repositorio contiene los scripts necesarios para optimizar un modelo de LightGBM para clasificación binaria usando optimización bayesiana. Hay dos pasos principales, cada uno en su propio archivo de script. Además, si prefieres saltar el primer paso, que implica el proceso de optimización que puede tomar tiempo, puedes utilizar los mejores hiperparámetros precomputados en el archivo HT4220_mejor.txt.

Archivos:

1. step_1_lightgbm_binaria_BO.r:

- **Propósito:** Este script realiza la optimización bayesiana para encontrar los mejores hiperparámetros para un modelo de clasificación binaria de LightGBM. Afina parámetros como `learning_rate`, `num_leaves`, `bagging_fraction`, entre otros.
- **Cómo ejecutarlo:**
 - Carga tu dataset y ejecuta este script para comenzar el proceso de optimización.
 - La salida incluirá los mejores hiperparámetros, que se guardarán automáticamente en la carpeta.
- **Nota:** Este paso puede tomar tiempo dependiendo del tamaño de tu dataset y la complejidad del modelo.

2. step_2_lightgbm_final.r:

- **Propósito:** Este script entrena el modelo final de LightGBM utilizando los mejores hiperparámetros obtenidos en step_1 o los hiperparámetros precomputados en el archivo HT4220_mejor.txt.
- **Cómo ejecutarlo:**
 - Asegúrate de que el dataset esté correctamente cargado.
 - Si ya completaste la optimización (step 1), el script usará los mejores hiperparámetros de ese proceso.
 - Si no has ejecutado el step 1, puedes usar los hiperparámetros almacenados en HT4220_mejor.txt para ahorrar tiempo.

3. HT4220_mejor.txt:

- **Descripción:** Este archivo contiene los mejores hiperparámetros obtenidos a través de la optimización bayesiana en el dataset. Puede ser

utilizado para saltar el primer paso y proceder directamente al entrenamiento del modelo en `step_2_lightgbm_final.r`.

- **Columnas clave:**
 - `objective`, `metric`, `learning_rate`, `num_leaves`, `feature_fraction`, etc., son algunos de los hiperparámetros ajustados.
 - `ganancia` se refiere a la ganancia del modelo, y `envios` indica el número de iteraciones.

Cómo saltar el Step 1:

Si prefieres evitar el proceso de optimización, simplemente carga los mejores hiperparámetros desde `HT4220_mejor.txt` en el script `step_2_lightgbm_final.r`. El script está diseñado para aceptar estos valores y entrenar el modelo directamente sin la necesidad de ejecutar la optimización bayesiana, que es más intensiva en tiempo.

Instrucciones de Ejecución:

1. Ejecuta los scripts R usando tu entorno de R preferido o directamente desde tu carpeta de Git.
2. **Nota Importante: Ten cuidado con las rutas de los archivos.** Asegúrate de que las rutas a tu dataset y otros archivos estén modificadas correctamente para que coincidan con la ubicación en tu ordenador.
3. Si no quieres ejecutar el step 1, carga los hiperparámetros desde `HT4220_mejor.txt` y continúa con el step 2.

Notas adicionales:

- Los hiperparámetros precomputados en `HT4220_mejor.txt` han sido optimizados específicamente para este dataset. Para datasets diferentes, se recomienda volver a ejecutar la optimización.
- Asegúrate de tener instaladas todas las bibliotecas necesarias (por ejemplo, `lightgbm`, `bayesianoptimization`) antes de ejecutar los scripts.