Parcial 2
Predicción de Ingresos con Datos del Censo

Contexto

En este proyecto analizarás el dataset Adult Census Income del repositorio UCI Machine Learning, que contiene datos del censo estadounidense de 1994. El objetivo es desarrollar una red neuronal que prediga si una persona gana más de $50,000 anuales basándose en características demográficas y socioeconómicas.

Reglas Generales

- Puedes utilizar recursos de internet como referencia, pero debes **comprender** el código, no solo copiarlo
- Utiliza **PyTorch** obligatoriamente para implementar redes neuronales
- Documenta tu código y decisiones técnicas

1 Recolección y procesamiento de datos (10%)

1. Diríjase a la URL https://archive.ics.uci.edu/dataset/2/adult y descargue el dataset. cargue los archivos importantes (datos de entrenamiento y de prueba) en Google Colab.
2. Con los datos de **prueba**, haga un split 50/50 para crear los datos de **validación**.
3. Realice Exploratory Data Analysis (EDA).
4. Haga un procesamiento de los datos para sus modelos. Cuidado con el **Data Leakage**

2 Desarrollo de algoritmos (50%)

2.1 Modelo Baseline (10%)

1. Realice un modelo de regresión logística y entrénelo.
2. Obtenga las métricas (las que se usan para un problema de clasificación binario) de entrenamiento/validación/prueba.

2.2 Modelo de Redes Neuronales (40%)

1. Cree la arquitectura base de la red (MLP)
2. Defina la función de pérdida y el optimizador.
3. Cree un loop de entrenamiento que incluya la validación.
4. Realice al menos 5 experimentos con distintas configuraciones de hiperparámetros. Utilice la GPU. Sea generoso con el número de épocas, capas y neuronas. Haga las gráficas de pérdida vs épocas. Analice las gráficas y detecte si hay overfitting/underfitting.

5. Investigue en internet e implemente las siguientes técnicas: `Dropout` y `EarlyStopping`. Pista: `Dropout` se implementa dentro de la arquitectura base de la red. `EarlyStopping` se usa dentro del loop de entrenamiento.

6. Vuelva a realizar (4) y obtenga el mejor MLP con regularización.

7. Obtenga las métricas (las que se usan para un problema de clasificación binario) de entrenamiento/validación/prueba.

3 Reporte y GitHub (20%)

1. Cree un reporte con lo siguiente:

   - Decisiones del procesamiento de datos (Si usaron todas las características, si crearon unas nuevas, el tipo de transformaciones que hicieron, etc.)
   - Hiperparámetros del mejor experimento de MLP. Apóyese de las gráficas generadas también. Compare los mejores MLP sin regularización y con regularización.
   - Compare el mejor MLP y la regresión lineal a partir de sus métricas. Interprete los resultados.

2. Cargue el reporte y el notebook a GitHub.

3. El notebook se debe correr de inicio a fin y al finalizar debe dar los resultados del mejor MLP. Si los resultados no tienen nada de similitud, se les bajará puntos.

4. Si el notebook genera algún error de inicio a fin y no se pueden generar los resultados del reporte, se bajarán puntos.