

Entrega 0: Proyecto para ramo INF395

Introducción a las Redes Neuronales y Deep Learning

Integrantes:

- Bruno Morici Rol: 202373555-8
- Juan Pablo Fuenzalida Rol: 202373102-1

Ramo:

- Sigla: INF395
- Semestre: 2025-2
- Profesor: Alejandro Veloz

Título del tema

Clasificación de eventos de colisión electrónica en CERN usando aprendizaje profundo

Tipo de problema

Aprendizaje supervisado – clasificación/regresión (predicción de masa invariante, clasificación de eventos físicos)

Descripción

Se dispone de un conjunto de datos de colisiones de dos electrones (dielectron events) con características como energía, momento, cargas, ángulos y masa invariante. El dataset contiene alrededor de 100 000 eventos en el rango de masa invariante entre 2 y 110 GeV. El objetivo es utilizar técnicas de redes neuronales (por ejemplo, redes densas, CNN o GNN) para predecir la masa invariante o clasificar tipos de eventos a partir de las características disponibles.

Problema a abordar por el grupo

Nuestro objetivo es implementar y evaluar modelos de aprendizaje profundo que puedan:

- Predecir la masa invariante (regresión) con mínima pérdida (ej. MSE).
- (Opcional) Clasificar eventos que podrían indicar procesos físicos relevantes.
- Comparar resultados con implementaciones existentes como la predicción de masa usando ANN (ver repositorio en GitHub).

Punto de partida

Existe un repositorio público en GitHub que contiene el dataset (`dielectron.csv`) y un notebook (`prediction.ipynb`) con una predicción inicial mediante redes neuronales: <https://github.com/IsaíasN/cern-electron-collision-data-m-prediction>. Nuestro trabajo se diferenciará al explorar:

- Arquitecturas más sofisticadas (CNN sobre representación gráfica de los eventos, o GNN).
- Ajuste de hiperparámetros y comparación de rendimiento.
- Análisis de importancia de características (feature importance).

Dataset(s)

El dataset contiene 100 000 eventos de colisión electrónica con las siguientes variables:

- Energías (E1, E2)

- Componentes del momento ($px_1, py_1, pz_1, px_2, py_2, pz_2$)
- Momento transverso (pt_1, pt_2)
- Pseudorapidez (η_1, η_2)
- Ángulo phi (ϕ_1, ϕ_2)
- Carga (Q_1, Q_2)
- Masa invariante (M)

Disponible en Kaggle: <https://www.kaggle.com/datasets/fedesoriano/cern-electron-collision-data>
 También hay un notebook de Kaggle que explora algoritmos de árbol de decisión: <https://www.kaggle.com/code/mariyagoliyad/dt-algorithms-for-cern-electron-collision-data>.

Artículos relacionados

- Andrews et al. (2018): Aplicación de clasificadores end-to-end basados en imagen para distinguir eventos de colisión en el LHC usando datos CMS (<https://arxiv.org/abs/1807.11916>).
- Fernández Madrazo et al. (2017): Aplicación de redes convolucionales para clasificar colisiones en física de alta energía transformando datos físicos en imágenes (<https://arxiv.org/abs/1708.07034>).
- Mikołaj Kita et al. (2024): Uso de modelos de difusión generativa para acelerar simulaciones de colisiones en experimentos de CERN (<https://arxiv.org/abs/2406.03233>).
- Verdone et al. (2024): Mejora del análisis de colisiones en física de partículas con Graph Neural Networks y técnicas de atribución de datos (<https://arxiv.org/abs/2407.14859>).
- Tesis de identificación de partículas en ALICE usando ML y modelos generativos (<https://alice-collaboration.web.cern.ch/node/35772>).