Module Project | Deep Learning | Cohort 11

Objective

Design, train, and evaluate a deep learning model using tabular data. Your goal is to apply what you've learned about model architecture, optimization, and regularization to solve a real-world binary classification task.

Dataset

Tabular — Higgs Boson Signal Classification

- **Dataset**: Kaggle Higgs Boson Machine Learning Challenge
- **Task**: Predict whether a particle collision event corresponds to a Higgs boson signal or just background noise.
- **Features**: 30 anonymized real-valued features derived from particle physics measurements.

Tasks to Perform

1. Data Preparation

- Normalize continuous features
- Handle any missing or zero-filled values (if present)
- Train/validation/test split (e.g., 70/15/15)

2. Model Building

- Define architecture
  - Hidden layers, units, activation functions
  - Output layer with sigmoid activation
  - Loss function: binary crossentropy
- Set optimizer, learning rate, and epochs
- Apply regularization techniques (Dropout, BatchNorm, etc.)
- Use callbacks: EarlyStopping, ReduceLROnPlateau
- Try a deeper architecture or a different optimizer (RMSprop, AdamW)

3. Model Evaluation

- Plot training and validation accuracy/ loss curves
- Report performance on dataset using:
  - accuracy
  - precision, recall and F-1 score
  - Confusion matrix and ROC-AUC curve
- Compare performance with a tree-based baseline (e.g., XGBoost)

4. Reflection

Include short answers in your notebook:

- How did model depth and activation affect performance?
- What helped mitigate overfitting?

- How did the learning rate and optimizer affect convergence?
- What would you improve with more time or compute?