Report Submission

Title: Tabular Data Classification - Titanic Survival Prediction

Name: Shaik Farhan

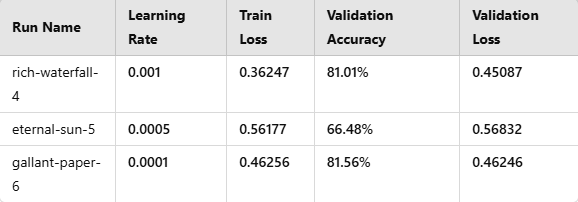
College: Vellore institute of technology

Reason why I chose this: I selected the Titanic Survival Prediction project because of my prior experience with tabular data classification, which I studied in my 5th semester. This familiarity allows me to effectively preprocess the dataset and apply machine learning models.

By choosing this task, I could focus more on model optimization and experiment tracking using Weights & Biases (W&B) rather than spending excessive time on data preprocessing and complex feature engineering.

Algorithm Analysis:

* **Objective:** Predict passenger survival based on features like age, sex, and class.
* **Dataset:** Titanic dataset from Kaggle(https://www.kaggle.com/c/titanic/data)
* **Model:** Feedforward Neural Network (FNN)
* **Tools Used:** Python, PyTorch, W&B for experiment tracking
* **Data Preprocessing:**
  + Handling missing values (e.g., Age, Embarke
  + Encoding categorical variables (Sex, Embarked)
  + Feature scaling (StandardScaler)
* **Neural Network Architecture:**
  + Input → Hidden Layers (16, 8 neurons) → Output (Sigmoid)
  + Activation: ReLU (Hidden), Sigmoid (Output)
  + Loss Function: Binary Cross-Entropy
* **Hyperparameters Tested:**
  + Learning Rate: 0.001, 0.0005, 0.0001
  + Optimizers: Adam, SGD



**Experiment 1 - Learning Rate = 0.001**

* **Observations:**
  + Low training loss (0.36247) suggests good convergence.
  + Validation accuracy of **81.01%** is relatively high.
  + Validation loss (0.45087) is stable, indicating minimal overfitting.
* **Conclusion:** Balanced performance, effective learning.

**Experiment 2 - Learning Rate = 0.0005**

* **Observations:**
  + High training loss (0.56177) indicates slow learning.
  + Validation accuracy is **only 66.48%**, much lower than others.
  + Validation loss (0.56832) remains high, suggesting **underfitting**.
* **Conclusion:** Learning rate too low, model struggles to learn effectively

**Experiment 3 - Learning Rate = 0.0001**

* **Observations:**
  + Training loss (0.46256) is higher than **0.001 LR**, but better than **0.0005 LR**.
  + Validation accuracy of **81.56%**, the highest among all experiments.
  + Validation loss (0.46246) is stable, indicating good generalization.
* **Conclusion:** Best validation accuracy, optimal learning rate balance.

Conclusions:

* **Best Learning Rate:** 0.0001 achieved the highest validation accuracy (81.56%).
* **Underfitting Issue:** LR = 0.0005 resulted in the worst performance.
* **Training vs. Validation Tradeoff:** Lower LR (0.0001) gave slightly better generalization.

**Wandb links:**

[**Click to go to wandb**](https://api.wandb.ai/links/shaikfarhan9492-vellore-institute-of-technology/wnk75xc9)

Thank you