NLP 243 Homework 1 – Relation Extraction from Natural Language

Submitted by: Bab Jan

1. Understanding the Task and Dataset

The objective of this assignment is to build a model that identifies the type of information a user is seeking in a natural language query. This is formulated as a **multi-label classification** problem because a single query may refer to multiple relation types (or none).

The dataset contains user queries (train.csv, val.csv, test.csv) along with corresponding relation labels. There are 19 possible relation labels, listed in all_labels.csv. I used a **Bag-of-Words (BoW)** representation, created via vectorizer.joblib, to convert queries into fixed-size feature vectors. This approach simplifies input representation while providing reasonable baseline performance.

2. Model Architecture

Implemented a **multi-layer perceptron (MLP)** with the following structure:

- Input layer: 942 features (BoW vocabulary size)
- Hidden layers: 3 fully connected layers with ReLU activations
- Dropout: Applied with 0.3 probability to prevent overfitting
- Output layer: 19 units, corresponding to the relation labels, with **sigmoid activation** for multi-label prediction

The model is trained with **binary cross-entropy loss**, suitable for multi-label classification. Training is performed using the **Adam optimizer**.

Implemented **early stopping** with patience of 5 epochs to prevent overfitting, monitoring the weighted F1 score on the validation set.

3. Experiments and Hyperparameter Variations

We experimented with three main hyperparameters, exploring 9 different combinations:

Experiment	Hidden Layers	Learning Rate	Batch Size
1	2	0.001	32
2	2	0.001	64
3	2	0.0005	64
4	3	0.001	32
5	3	0.001	64

Experiment	Hidden Layers	Learning Rate	Batch Size
6	3	0.0005	64
7	4	0.001	32
8	4	0.001	64
9	4	0.0005	64

Observations:

- Increasing hidden layers improved performance up to 3 layers; 4 layers did not significantly improve F1.
- Larger batch sizes (64) accelerated convergence without hurting performance.
- Lower learning rates (0.0005) slowed training but slightly stabilized final loss.

Best model: 3 hidden layers, learning rate 0.001, batch size 64. Achieved **weighted F1 = 0.8983** on the validation set and **0.8957** on the test set.

4. Training and Validation Loss

- Loss decreased steadily on both training and validation sets.
- Early stopping triggered at epoch 29, preventing overfitting.
- Validation weighted F1 improved consistently, demonstrating good generalization.

5. Results and Findings

- Weighted F1 on test set: **0.8957**
- Class-wise performance shows strong results for majority classes, with lower scores for classes with zero support.
- Dropout and early stopping prevented overfitting, especially for smaller classes.
- Hyperparameter tuning had significant impact on convergence speed and final F1 scores.

6. Conclusion

This assignment demonstrated the effectiveness of a simple BoW + MLP model for multi-label relation extraction. By varying model depth, learning rate, and batch size, I identified a configuration that maximizes weighted F1 while preventing overfitting. The final model is robust and performs well across the validation and test datasets, achieving nearly 0.90 weighted F1.