Comprehensive Report for HW5, HW6, HW7, and HW8

1. Models Summary

This section provides an overview of the models and approaches used throughout HW5, HW6, HW7, and HW8, along with key hyperparameters.

HW5: Multilabel Classification with Feedforward Neural Networks

- Approach: Developed a feedforward neural network from scratch for multilabel emotion detection.
- Key Hyperparameters:
  - Learning rate: 0.001
  - Optimizer: Adam
  - Activation Function: ReLU
  - Epochs: 5
  - Loss Function: Binary Cross entropy
- Dataset: Processed the Kaggle emotion dataset for multilabel classification.
- Performance:
  - Accuracy and F1-Score metrics showed reasonable performance for foundational models.

HW6: Transformer-Based Models (Encoder-Only)

Models Used:

1. RoBERTa Base:
   - Pretrained encoder-only model for emotion detection.
   - Key Hyperparameters:
     - Batch size: 8
     - Learning rate: 5e-5
     - Scheduler: Linear decay with warmup
     - Epochs: 3
2. DistilBERT:
   - Lightweight model for improved efficiency.
   - Key Hyperparameters:
     - Batch size: 8
     - Learning rate: 5e-5

- Epochs: 3

3. DistilRoBERTa Base:

- Smaller RoBERTa variant for faster experimentation.

- Key Hyperparameters:

- Batch size: 8

- Learning rate: 5e-5

- Epochs: 3

- Special Features:

- Addressed class imbalance with weighted loss.

- Consistent evaluation metrics (F1-Score, Accuracy) for fair comparison.

HW7: Sentence Similarity with Embedding Models

Models Used:

1. LLaMA (meta-llama/Llama-3.2-1B):

- Applied for understanding sentence-level similarity using fine-tuned models.

- Key Hyperparameters:

- Embedding dimensions: Fixed at 768

- Optimizer: AdamW

- Epochs: 3

- Batch size: 16

- Learning rate: 1e-5

2. Gemma (google/gemma-2-2b):

- Specialized for embedding tasks, particularly on sentence-to-sentence evaluations.

- Key Hyperparameters:

- Embedding dimensions: Fixed at 768

- Optimizer: AdamW

- Epochs: 3

- Batch size: 4

- Learning rate: 1e-5

3. MTEB (Massive Text Embedding Benchmark) (intfloat/e5-mistral-7b-instruct):

- Evaluated embeddings for zero-shot classification and semantic similarity tasks.

- Key Hyperparameters:

- Embedding dimensions: Fixed at 768
- Optimizer: AdamW
- Epochs: 5
- Batch size: 32
- Learning rate: 5e-6

HW8: Zero-Shot and Instruction-Tuned Models

Models Used:

1. Base Model:
   - Evaluated without task-specific fine-tuning.

- Key Hyperparameters:
  - Embedding dimensions: Fixed at 768
  - Optimizer: AdamW
  - Epochs: 3
  - Batch size: 16
  - Learning rate: 1e-5

2. Instruction-Tuned Model:
   - Fine-tuned with instruction datasets for better task adaptability.

- Key Hyperparameters:
  - Embedding dimensions: Fixed at 768
  - Optimizer: AdamW
  - Epochs: 3
  - Batch size: 16
  - Learning rate: 1e-5

3. Zero-Shot Classification:
   - Leveraged models like BART-large-mnli for direct emotion detection without additional fine-tuning.

- Key Features:
  - Enabled batch processing for efficiency.
  - Applied class weights to handle label imbalance.

---

2. Performance Analysis

This section compares the performance of models across the assignments.

HW5

- Strengths:
  - Simple architecture ensured interpretability.
  - Baseline for multilabel classification.
- Weaknesses:
  - Struggled with high-dimensional input.
  - Limited capacity to capture complex semantic relationships.
  - Low F1 – 0.1361 and low accuracy - 0.056921

HW6

- RoBERTa Base:
  - Strengths: Best performance in terms of F1-Score (macro average: 0.5897).
  - Weaknesses: Computationally expensive.
- DistilBERT:
  - Strengths: Faster training and inference times.
  - Weaknesses: Slightly lower F1-Score (macro average: 0.5806).
- DistilRoBERTa Base:
  - Strengths: Balance between speed and performance.
  - Weaknesses: Slightly less robust than RoBERTa Base. (F1 = 0.5703)

HW7

- LLaMA:
  - Strengths: Strong semantic understanding of sentence similarity.
  - Weaknesses: Resource-intensive. (F1=0.4806)
- Gemma:
  - Strengths: Lightweight and scalable. (F1=0.5377)
  - Weaknesses: Lower accuracy in edge cases.
- MTEB:
  - Strengths: Generalizability across benchmarks.

- Weaknesses: Significant preprocessing is required for diverse tasks. (F1=0.2663)

HW8

- Base Model:
  - Strengths: Provides a baseline for zero-shot tasks.
  - Weaknesses: Limited performance without fine-tuning. (F1=0.4797)
- Instruction-Tuned Model:
  - Strengths: Significant improvement in adaptability and F1-Score (0.5271).
  - Weaknesses: Requires careful curation of instruction datasets.
- Zero-Shot Classification:
  - Strengths: Efficient for unseen tasks, particularly with BART-large-mnli.
  - Weaknesses: Accuracy can drop for nuanced examples. (very low F1 score)

---

3. Future Approach and Lessons Learned

Future Approaches

1. Improved Data Augmentation:
   - Use advanced augmentation techniques to address class imbalance and increase dataset diversity.
2. Alternative Architectures:
   - Explore hybrid models combining encoder-decoder setups for contextual understanding.
3. Unsupervised Pretraining:
   - Train domain-specific embeddings for better alignment with the target tasks.
4. Scalable Zero-Shot Techniques:
   - Experiment with newer models like T5 or GPT variants for zero-shot and few-shot tasks.

Lessons Learned

- Model Selection:
  - Encoder-only models excel at structured tasks like classification.
  - Decoder or instruction-tuned models are versatile but resource-intensive.
- Preprocessing:
  - Consistent preprocessing improves robustness across datasets.

- Evaluation Metrics:
  - F1-Score and recall are crucial for understanding multilabel performance.
- Efficiency vs. Accuracy Trade-off:
  - Lighter models like DistilBERT provide faster results but may compromise accuracy.

---

4. Public Weights & Biases Link

HW6

RoBERTa Base- https://wandb.ai/pxy230011-the-university-of-texas-at-dallas/Exp1?nw=nwuserpxy230011

DistilBERT- https://wandb.ai/pxy230011-the-university-of-texas-at-dallas/Exp2?nw=nwuserpxy230011

Similar-Sized Model- https://wandb.ai/pxy230011-the-university-of-texas-at-dallas/Exp3?nw=nwuserpxy230011

HW7

google/gemma- https://wandb.ai/pxy230011-the-university-of-texas-at-dallas/gemma?nw=nwuserpxy230011

meta/Llama- https://wandb.ai/pxy230011-the-university-of-texas-at-dallas/LLama?nw=nwuserpxy230011

MTEB- https://wandb.ai/pxy230011-the-university-of-texas-at-dallas/MTEB?nw=nwuserpxy230011

HW8

Base- https://wandb.ai/pxy230011-the-university-of-texas-at-dallas/Base?nw=nwuserpxy230011

Instruction-tuned- https://wandb.ai/pxy230011-the-university-of-texas-at-dallas/Instruction-tuned?nw=nwuserpxy230011

Zero-Shot- https://wandb.ai/pxy230011-the-university-of-texas-at-dallas/zero-shot?nw=nwuserpxy230011