CS771 Mini-Project 2

Avni Maheshwari
Roll No. 210241

Manjusree Nayak
Roll No. 210587

Priya
Roll No. 210771

Shruthi Munnur
Roll No. 211015

Snehal Shridhar Kane
Roll No. 211044

1 Problem 1

1.1 Task 1

The task involves processing datasets D_1 to D_{10}, each having the same input distribution, to train models iteratively while retaining performance on previous datasets. The code implements this using the following approach:

1. **Feature Extraction**:
   - A pre-trained MobileNetV2 model is employed for feature extraction. The CIFAR-10 images are resized from 32×32 to 96×96, as the model expects higher resolution inputs.
   - Features are normalized to improve cosine similarity computations, ensuring robust comparisons across datasets.
2. **LwP Classifier**:
   - The *Learn with Prototypes (LwP)* classifier initializes class prototypes based on labeled data in D_1. The prototypes represent the average feature vector for each class.
   - The classifier predicts labels by computing normalized cosine similarity scores, treating these scores as pseudo-probabilities for classification.
3. **Iterative Updates**:
   - The classifier, trained on D_1, predicts labels for the next dataset D_2. Using these pseudo-labels, the model prototypes are updated incrementally with new feature data, while freezing earlier datasets' influence.
   - This process repeats iteratively for datasets D_2 through D_{10}, ensuring the prototype space evolves while preserving information from prior datasets.
4. **Evaluation**:
   - Each model f_i is evaluated on its corresponding held-out dataset $\hat{D}_i$ and all previous datasets $\hat{D}_j, j < i$, using accuracy as the primary metric. The goal is to minimize performance degradation on earlier datasets as the model adapts to new data.

Models Used	**Accuracy on Test Data**									
	$\hat{D}_1$	$\hat{D}_2$	$\hat{D}_3$	$\hat{D}_4$	$\hat{D}_5$	$\hat{D}_6$	$\hat{D}_7$	$\hat{D}_8$	$\hat{D}_9$	$\hat{D}_{10}$
F_1	0.7504									
F_2	0.7504	0.7488								
F_3	0.7476	0.7484	0.7304							
F_4	0.7484	0.748	0.7304	0.7468						
F_5	0.7488	0.7476	0.7292	0.7468	0.764					
F_6	0.7472	0.748	0.7308	0.7452	0.7624	0.7612				
F_7	0.7468	0.7472	0.7302	0.7456	0.762	0.76	0.7328			
F_8	0.7456	0.7472	0.7296	0.7456	0.7608	0.7596	0.7328	0.742		
F_9	0.7464	0.7472	0.728	0.7448	0.762	0.7592	0.7324	0.7408	0.7332	
F_{10}	0.7464	0.7476	0.7292	0.7448	0.76	0.7588	0.7336	0.7412	0.7328	0.7492

Table 1: Table Showing the Models Used and their Accuracy on Test Data

1.2 Task 2

Starting with the model f_{10} from Task 1, the goal is to adapt the classifier for datasets D_{11} to D_{20}, where the input distributions vary. The code adapts the previous methodology with considerations for distribution shifts:

1. **Feature Alignment**:
   - Features extracted using MobileNetV2 undergo the same preprocessing as in Task 1. The consistency in feature extraction helps mitigate domain shifts during updates.
2. **Prototype Adaptation**:
   - The LwP classifier is fine-tuned iteratively by aligning class prototypes to new datasets. High-confidence predictions guide pseudo-labeling, with weights emphasizing similarities between new and previous distributions.
   - To address distribution shifts, prototypes are updated with a regularization strategy that balances new data's influence against the stability of existing prototypes.
3. **Regularization Against Drift**:
   - The prototypes for classes in earlier datasets are preserved using cosine similarity regularization. This ensures that while the model adapts to new distributions, its performance on previous datasets remains consistent.
4. **Evaluation Matrix**:
   - Each model f_{i+10} (for $i \in \{11, \ldots, 20\}$) is evaluated across all 20 held-out datasets. The evaluation matrix captures the classifier's adaptability and stability across domains.

This approach combines a feature-extraction pipeline with iterative updates and regularization, effectively handling both uniform and non-uniform data distributions.

Models Used	Accuracy on Test Data																			
	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	D_{11}	D_{12}	D_{13}	D_{14}	D_{15}	D_{16}	D_{17}	D_{18}	D_{19}	D_{20}
F_{11}	0.746	0.748	0.729	0.745	0.76	0.759	0.734	0.741	0.733	0.749	0.573									
F_{12}	0.746	0.748	0.729	0.745	0.76	0.759	0.733	0.741	0.733	0.749	0.573	0.498								
F_{13}	0.746	0.748	0.729	0.745	0.76	0.759	0.734	0.741	0.733	0.749	0.573	0.498	0.676							
F_{14}	0.746	0.745	0.729	0.745	0.76	0.759	0.734	0.741	0.733	0.749	0.573	0.498	0.676	0.626						
F_{15}	0.746	0.747	0.729	0.745	0.76	0.759	0.734	0.741	0.733	0.749	0.573	0.498	0.676	0.626	0.75					
F_{16}	0.746	0.748	0.729	0.745	0.76	0.759	0.734	0.741	0.733	0.749	0.573	0.498	0.676	0.626	0.75	0.596				
F_{17}	0.746	0.747	0.729	0.745	0.76	0.759	0.734	0.741	0.733	0.749	0.573	0.498	0.676	0.626	0.75	0.596	0.613			
F_{18}	0.746	0.748	0.729	0.745	0.76	0.759	0.734	0.741	0.733	0.749	0.573	0.498	0.676	0.626	0.75	0.596	0.613	0.676		
F_{19}	0.746	0.748	0.729	0.745	0.76	0.759	0.734	0.741	0.733	0.749	0.573	0.498	0.676	0.626	0.75	0.596	0.613	0.676	0.504	
F_{20}	0.746	0.748	0.729	0.745	0.76	0.759	0.734	0.741	0.733	0.749	0.573	0.498	0.676	0.626	0.75	0.596	0.613	0.676	0.504	0.717

Table 2: Table Showing the Models Used and their Accuracy on Test Data

2 Problem 2

Video Link: Lifelong Domain Adaptation via Consolidated Internal Distribution.