Assignment - 2

Name: Vivek Sapkal *Roll No.: B22AI066*

Task 0: Evaluation of Algorithms on IAPRTC dataset

Introduction

This report evaluates six machine learning algorithms for multi-label classification on the IAPRTC-12 dataset, which contains 17,665 training and 1,962 test samples with 2,048-dimensional visual features and 291 possible labels. The evaluated models include:

1. Linear SVM
2. Logistic Regression
3. Softmax Regression
4. Decision Tree
5. Weighted kNN
6. Ensemble Model

Methodology

Dataset Preprocessing

1. Data Loading:
    - The dataset was loaded from .mat files, with features (I_x_tr, I_x_te) and labels (I_z_tr, I_z_te) standardized using StandardScaler to ensure feature comparability.
    - PCA reduced feature dimensions from 2,048 to 200 while preserving 72% of the variance, addressing computational complexity and the curse of dimensionality.

Model Implementations

All models were implemented using scikit-learn with a One-vs-Rest strategy to handle multi-label classification:

- Linear SVM: LinearSVC with early stopping (max_iter=1).
- Logistic Regression: L2 regularization and OneVsRestClassifier.

- Softmax Regression: Multinomial logistic regression with lbfgs solver.
- Decision Tree: DecisionTreeClassifier with depth=5 and Gini impurity.
- Weighted kNN: KNeighborsClassifier (k=5, distance weighting).
- Ensemble: Majority voting across all models (threshold ≥3 votes).

Evaluation Metrics

Models were assessed using:

- Accuracy, Precision (micro/macro), Recall (micro/macro)
- F1 Score (micro/macro), Hamming Loss, Training Time

Results

Performance Overview

Model	Accuracy	Precision (Micro)	Recall (Micro)	F1 (Micro)	Hamming Loss
Linear SVM	0.0183	0.7566	0.2255	0.3475	0.0164
Logistic Regression	0.0000	0.0704	0.8624	0.1302	0.2230
Softmax Regression	0.0000	0.0704	0.8624	0.1302	0.2230
Decision Tree	0.0102	0.5300	0.2440	0.3342	0.0188
Weighted kNN	0.0347	0.8521	0.3971	0.5417	0.0130
Ensemble	0.0255	0.6429	0.4830	0.5516	0.0152

Key Observations

1. Weighted kNN achieved the lowest Hamming Loss (0.013), indicating strong multi-label compatibility.
2. Logistic/Softmax Regression failed completely (0% accuracy), likely due to insufficient iterations (max_iter=1) and class imbalance.
3. Ensemble Model balanced performance with the highest recall (0.483) and competitive F1 scores.
4. Linear SVM showed high precision (0.7566) but poor recall, suggesting overfitting to frequent labels.

Challenges

1. Class Imbalance: With 291 labels, rare classes likely hindered logistic/softmax regression.
2. High Dimensionality: PCA mitigated this, but feature reduction might have discarded discriminative information.
3. Model-Specific Limitations:
    - Linear models (SVM, logistic regression) struggled with non-linear label correlations.
    - Decision Trees underperformed due to depth constraints (max_depth=5).
4. Computational Costs: Training time varied significantly, with kNN being resource-intensive due to PCA-transformed features.

Discussion

- Weighted kNN excelled by leveraging local feature relationships, critical for multi-label tasks.
- The Ensemble combined diverse model strengths, improving robustness (e.g., recall increased by ~20% over individual models).
- Micro vs. Macro Metrics: Micro-averaged metrics favored frequent labels, while macro-averaged scores highlighted class-wise imbalances.

References

1. Scikit-learn Developers. (2023). *Scikit-learn: Machine Learning in Python*. [Online]. Available: https://scikit-learn.org
2. IAPRTC-12 Dataset. (n.d.). *Image Annotation and Retrieval*. [Online]. Available: Dataset Link

Visualizations