

Intro to Machine Learning
HW07

Conclusion for Adaboost Algorithm:

- The AdaBoost classifier achieved a test accuracy of 65%, showing moderate effectiveness in classification.
- It successfully identified Category 1 with a significant number of true positives (10 instances).
- The model displayed confusion between Category 1 and 3, misclassifying Category 3 as Category 1 three times.
- There were no predictions for Category 2, indicating possible difficulties in recognizing this class or a class imbalance.
- Improvement could be sought through careful feature selection, parameter optimization, and addressing potential class imbalance.

Comparison with different:

1. **AdaBoost (HW8):**

- **Accuracy:** 65%.
- **Context:** Applied with a Decision Tree Classifier (max_depth=2) as the base estimator, learning rate of 0.01, and 100 estimators.

2. **Neural Network (HW5):**

- **Accuracy:** 75%.
- **Context:** Implemented using a two-layer neural network with ReLU and softmax activation functions. Although it had the highest accuracy, it faced limitations in predicting certain categories, with no correct predictions for category 2.

3. **Logistic Regression (HW6):**

- **Accuracy:** 80%.
- **Context:** Achieved higher accuracy than the neural network for classification tasks, suggesting its effectiveness in handling this particular dataset.

4. **Random Forest (HW7):**

- **Initial Accuracy:** 65%, similar to AdaBoost.
- **Enhanced Accuracy (with New Features):** 100%.
- **Context:** The initial performance was on par with AdaBoost. However, the accuracy improved dramatically to perfect accuracy with the addition of new features capturing the growth rate over time.

Comparative Analysis:

- **Neural Network vs. AdaBoost:** The neural network's accuracy was higher than AdaBoost but had significant limitations in predicting some categories.
- **AdaBoost vs. Logistic Regression:** AdaBoost's performance (65%) was lower compared to the Logistic Regression model (80%), suggesting that logistic regression was more suitable for this dataset.
- **AdaBoost vs. Random Forest (Initial):** Both models initially had the same level of accuracy. However, Random Forest's performance improved significantly with the addition of new features.

- **Random Forest (Enhanced) vs. AdaBoost:** Random Forest, with enhanced features, dramatically outperformed AdaBoost, achieving 100% accuracy.

Conclusion:

In this context, AdaBoost, while offering a balanced approach, was outperformed by both the Logistic Regression and the enhanced Random Forest model. The Logistic Regression model's higher accuracy (80%) indicates its effectiveness in linearly separating the classes in the dataset. The neural network, despite its higher accuracy, faced challenges in certain categories. The Random Forest model, particularly with the addition of new features, showed the highest potential in terms of adaptability and accuracy. This comparison highlights the importance of model selection based on the dataset's characteristics and the critical role of feature engineering in improving model performance.