HW8

AdaBoost Classifier:

```
...   Test Accuracy: 80.00%
```

Test Accuracy: 80%

- This is a simple and effective ensemble method that combines weak learners to create a strong classifier.
- Achieves reasonable accuracy, but its performance is surpassed by some other methods.

HW5 - Neural Network (TensorFlow and Keras):

Test Accuracy: 75%

- Architecture: 3 layers (6-6-3 nodes) with ReLU activation in the hidden layer, categorical cross-entropy loss, and Adam optimizer.
- Trained for 1000 epochs with a batch size of 32.
- Lower accuracy compared to AdaBoost, suggesting that the neural network may not be capturing the patterns effectively or needs further tuning.

HW6 - Logistic Regression:

Test Accuracy: 85%

- Stratified data split, multinomial multi-class setting, and hyperparameter tuning (max_iter=425).
- The simplicity of logistic regression works well for this task, outperforming both AdaBoost and the neural network in terms of accuracy.

HW7 - Random Forest:

Part 1: Test Accuracy: 75%

- Original dataset with normalized citation data.
- No additional feature engineering.
- Accuracy is lower than logistic regression and adaboost, indicating potential limitations in capturing patterns with the given features.

Part 2: Test Accuracy: 100%

- New features based on citation changes over the years.
- Exclusion of raw citation numbers.
- Improved accuracy, suggesting that the new features and the approach of focusing on changes contribute to better model performance.
- Random Forest is well-suited for handling non-linear relationships and benefits from the new feature set.

HW8

Comparison:

Random Forest, with appropriate feature engineering, achieves perfect accuracy in Part 2, emphasizing the importance of feature selection and model interpretability.

Logistic Regression outperforms other methods in terms of accuracy in this specific task.

AdaBoost and the neural network perform reasonably well but fall short of the accuracy achieved by logistic regression.

Several factors could contribute to AdaBoost not performing as well as other models in this specific classification task:

1. Weak Learner Choice: Adaboost's performance heavily relies on its weak learners, usually decision trees. If these learners fail to capture dataset patterns effectively, it can hinder Adaboost's accuracy.
2. Feature Engineering: The effectiveness of AdaBoost can be influenced by the quality of features. If the feature set lacks discriminatory power or if important features are missing, AdaBoost may not be able to build a robust model.
3. Underlying Relationships: Certain data relationships might be more suited for models like logistic regression, neural networks, or random forests, making Adaboost less effective in capturing the dataset's decision boundaries.

References:

https://scikit-learn.org/stable/modules/ensemble.html#adaboost