Report on Node2Vec and Logistic Regression for 7-Class Classification

Introduction:
This report presents findings from our utilization of the Node2Vec algorithm along with logistic regression for performing 7-class classification using the CORA graph dataset. Our goal was to generate node embeddings using Node2Vec and then employ these embeddings as features for logistic regression-based classification.

Results:

	precision	recall	f1-score	support
Case_Based	0.73	0.61	0.67	166
Genetic_Algorithms	0.88	0.87	0.88	210
Neural_Networks	0.70	0.85	0.77	366
Probabilistic_Methods	0.84	0.79	0.81	207
Reinforcement_Learning	0.86	0.71	0.78	128
Rule_Learning	0.69	0.56	0.62	98
Theory	0.62	0.63	0.63	174
accuracy			0.75	1349
macro avg	0.76	0.72	0.74	1349
weighted avg	0.76	0.75	0.75	1349

Discussion:
Our logistic regression model exhibited reasonable performance on the test set, achieving an accuracy of 75.24%. Analysis of precision, recall, and F1-score metrics revealed varying performance across different classes. For instance, Genetic_Algorithms displayed high precision, recall, and F1-score, indicating effective classification. However, Theory and Rule_Learning exhibited lower metrics, suggesting classification challenges for these classes.

Conclusion:
The integration of Node2Vec for generating node embeddings and logistic regression for classification proved effective for the 7-class classification task using the CORA graph dataset. Our attained accuracy and evaluation metrics offer insights into the model's performance across distinct classes.

Saikat Moi (20CS10050)

Report on Graph Convolutional Networks (GCN) for 7-Class Node Classification

Introduction:
This report presents the outcomes of applying the Graph Convolutional Network (GCN) architecture, as introduced by Kipf and Welling (2017), to the task of classifying nodes into 7 classes using the CORA graph dataset. Our aim was to implement the GCN model, train it on the dataset, and assess its performance in node classification.

Evaluation Results:

	precision	recall	f1-score	support
Theory	0.40	0.10	0.16	174
Reinforcement_Learning	0.56	0.04	0.07	128
Probabilistic_Methods	0.72	0.10	0.18	207
Case_Based	0.72	0.16	0.26	166
Genetic_Algorithms	0.82	0.18	0.29	210
Rule_Learning	0.37	0.07	0.12	98
Neural_Networks	0.30	0.95	0.45	366
accuracy			0.34	1349
macro avg	0.56	0.23	0.22	1349
weighted avg	0.54	0.34	0.26	1349

Discussion:
Analysis of precision, recall, and F1-score metrics revealed varying performance across different classes. For instance, Genetic_Algorithms,Case_Based, Probabilistic_Methods displayed high precision, recall, and F1-score, indicating effective classification. However, Neural_Networks,Theory and Rule_Learning exhibited lower metrics, suggesting classification challenges for these classes.

Comparison with Logistic Regression:
Comparing the GCN model with logistic regression from Part 1 reveals a significant performance gap. While logistic regression achieved an accuracy of 75.24% with balanced precision, recall, and F1-score metrics, the GCN model struggled with an accuracy of only 34%. Additionally, precision, recall, and F1-score were generally lower across classes.

Conclusion:
While the GCN model was successfully implemented and trained, its performance in node classification on the CORA graph dataset fell short of expectations. The comparison with logistic regression underscores the importance of selecting and fine-tuning models carefully for specific tasks and datasets.

Saikat Moi (20CS10050)