Home Work

Erik Norlin, 970807-9299, CID: norliner

June 21, 2023

Home work A

1a)

Number of training nodes: 140
Number of validation nodes: 500
Number of testing nodes: 1000

Supervised learning test accuracy: 0.59
Semi-supervised learning test accuracy: 0.815

Supervised learning was just standard dense layers (MLP) where as semi-supervised learning was a graph convolutional network. The MLP performed worse because it needs labels to be trained, and since the dataset is small the model suffers from overfitting. The graph neural network performed better because it uses all nodes for training even though most nodes are unlabelled. The idea is that the GCN clusters unlabelled nodes with labeled nodes that have similar features and graph structure, so unlabelled are trained on their features and graph structures. This ultimately leads to better performance since the GCN trains on more data. The great thing about this is that not as much labelled data is needed which consequently reduces training time and work to label data.

1c)

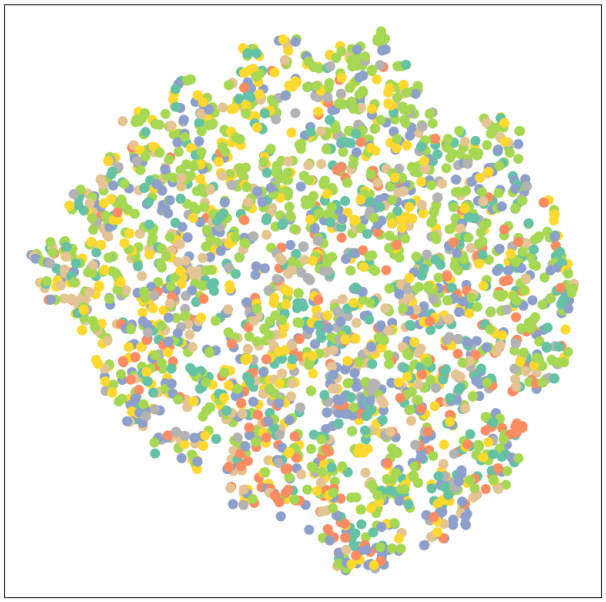

Figure 1: Untrained MLP.

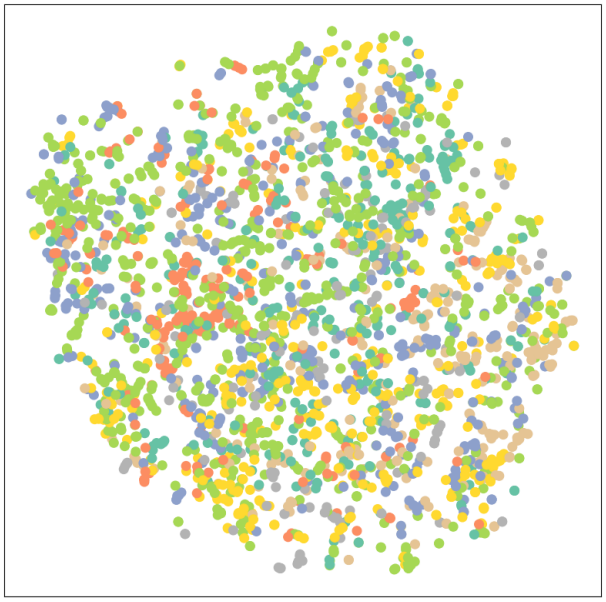

Figure 2: Untrained GCN

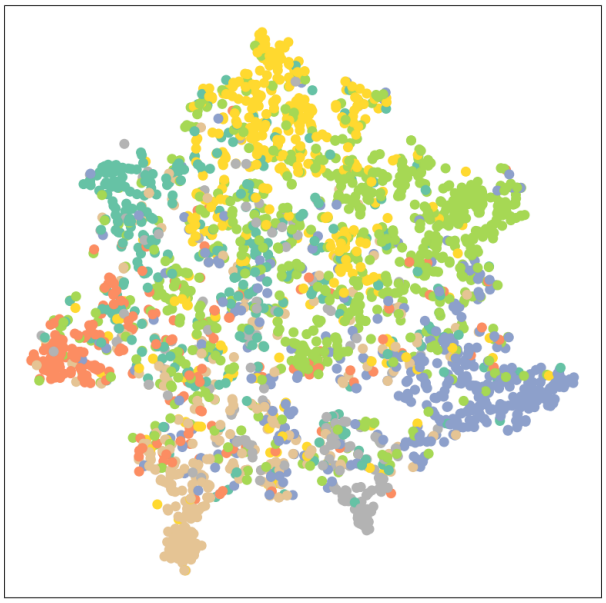

Figure 3: Trained MLP

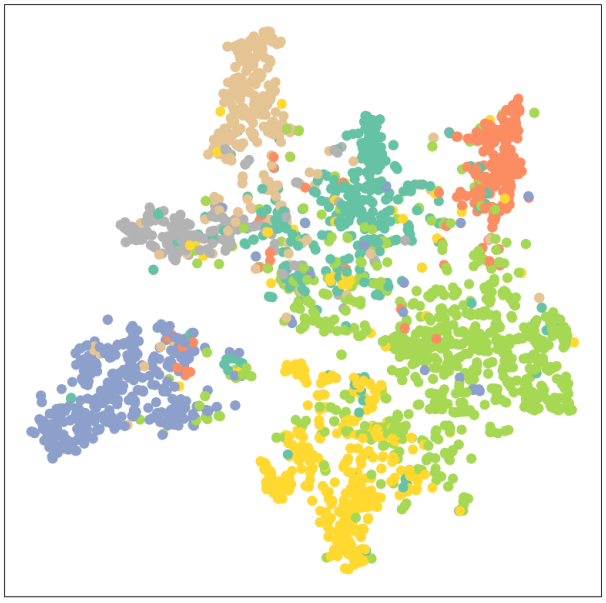

Figure 4: Trained GCN

2)

The reason for using the edge attributes as edge weight is that the edge attributes already has an underlying representation of the relationships between the nodes of the graph. This gives the network a better start and generally leads to better convergence.

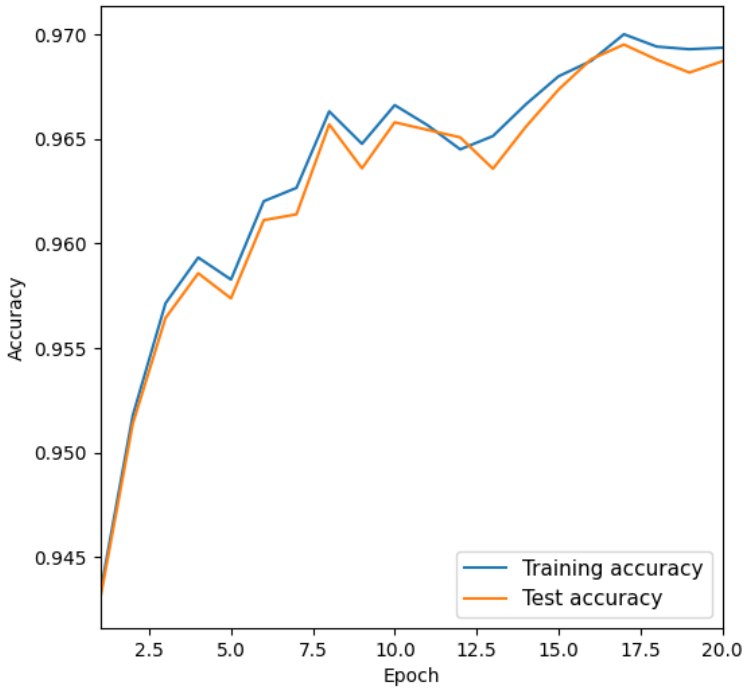

Figure 5: GCN test accuracy.