Assignment 3 Report

GMLFA (AI60007) - Autumn,2024 - IIT Kharagpur

Group 14

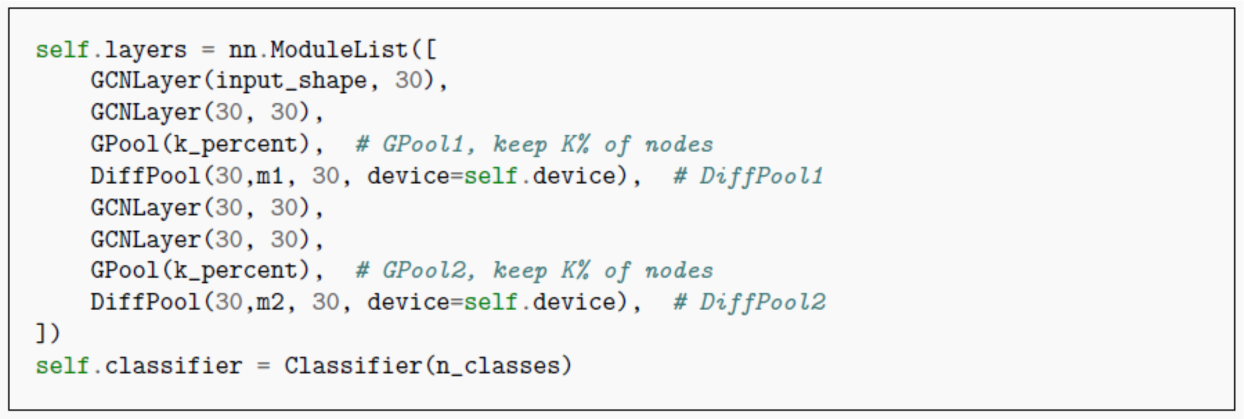
Vinayak Mali   Saurabh Jaiswal

24AI60R13   24AI60R46

October 17, 2024

Model Architecture

As instructed following is the model architecture



```
self.layers = nn.ModuleList([
    GCNLayer(input_shape, 30),
    GCNLayer(30, 30),
    GPool(k_percent),  # GPool1, keep K% of nodes
    DiffPool(30,m1, 30, device=self.device),  # DiffPool1
    GCNLayer(30, 30),
    GCNLayer(30, 30),
    GPool(k_percent),  # GPool2, keep K% of nodes
    DiffPool(30,m2, 30, device=self.device),  # DiffPool2
])
self.classifier = Classifier(n_classes)
```

The model architecture consists of a series of graph convolutional layers (GCNLayer) that process node features, followed by pooling layers (GPool) to retain a percentage of nodes (k percent). It incorporates differentiable pooling (DiffPool) to aggregate information and reduce dimensionality, ultimately feeding the output into a classifier to predict class labels.

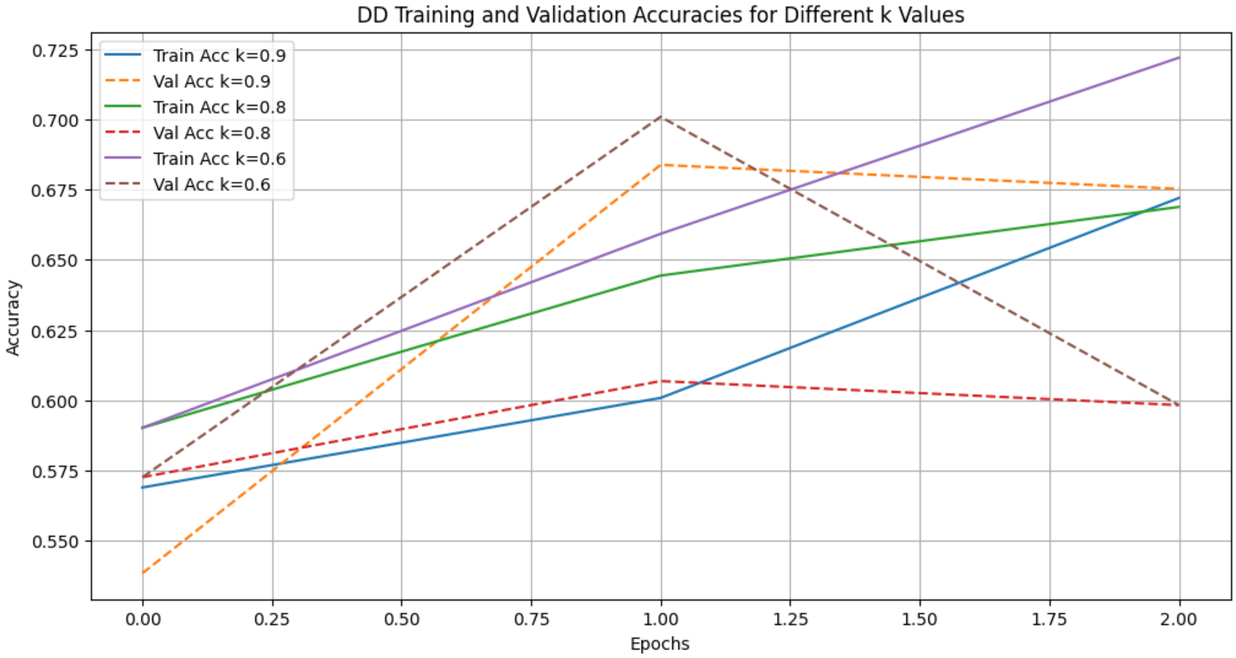
Analysis of D&D - Binary Classification Task

Final accuracy after <u>10</u> epoch with best_k 60% is : 77%

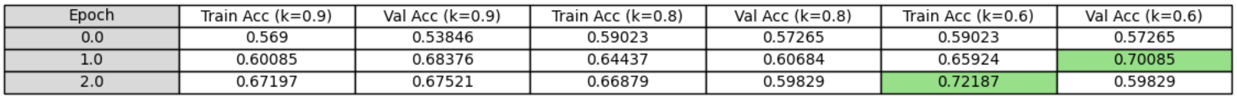
If number of epochs will be increased then it will improve the accuracy too.

We are going to analyze the loss and accuracy results on following points:
- cluster size in pooling layer i.e. m1 and m2 values
- By changing number of epochs (taking less number of epochs for visualization)

For epoch = 3 m1 = 6 and m2 = 3

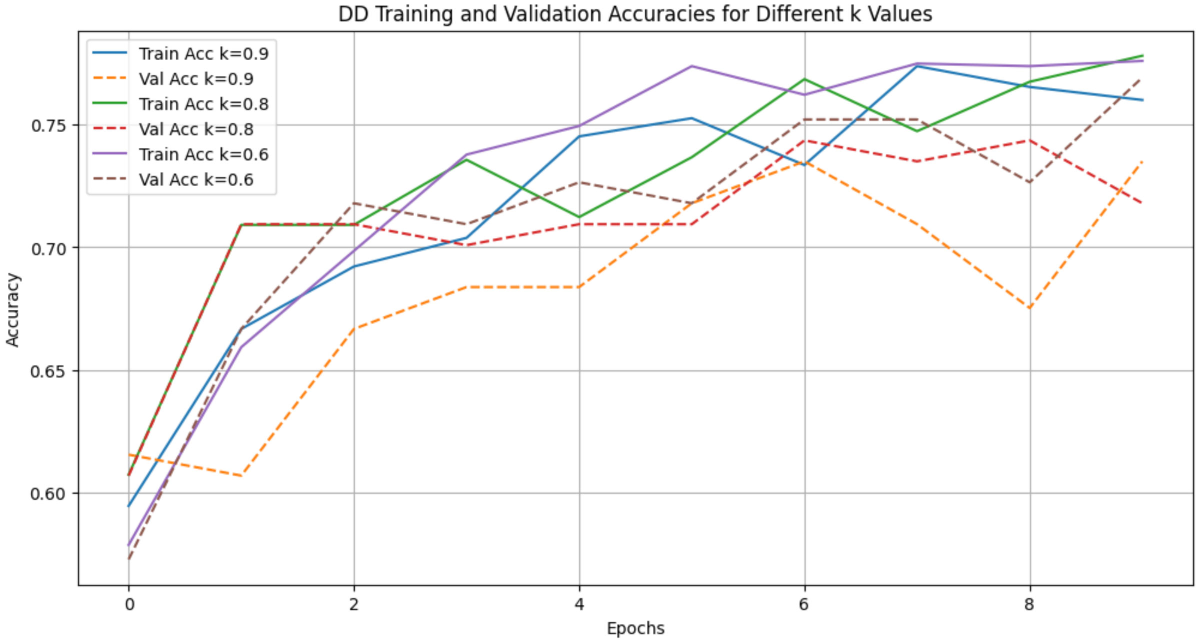


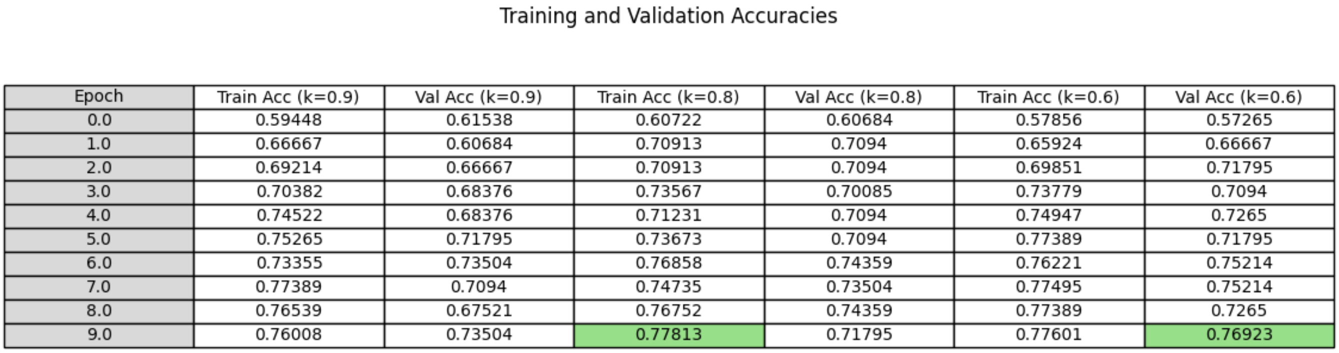
Training and Validation Accuracies



Epoch	Train Acc (k=0.9)	Val Acc (k=0.9)	Train Acc (k=0.8)	Val Acc (k=0.8)	Train Acc (k=0.6)	Val Acc (k=0.6)
0.0	0.569	0.53846	0.59023	0.57265	0.59023	0.57265
1.0	0.60085	0.68376	0.64437	0.60684	0.65924	0.70085
2.0	0.67197	0.67521	0.66879	0.59829	0.72187	0.59829

For epoch = 10 m1 = 6 and m2 = 3




Training and Validation Accuracies

Epoch	Train Acc (k=0.9)	Val Acc (k=0.9)	Train Acc (k=0.8)	Val Acc (k=0.8)	Train Acc (k=0.6)	Val Acc (k=0.6)
0.0	0.59448	0.61538	0.60722	0.60684	0.57856	0.57265
1.0	0.66667	0.60684	0.70913	0.7094	0.65924	0.66667
2.0	0.69214	0.66667	0.70913	0.7094	0.69851	0.71795
3.0	0.70382	0.68376	0.73567	0.70085	0.73779	0.7094
4.0	0.74522	0.68376	0.71231	0.7094	0.74947	0.7265
5.0	0.75265	0.71795	0.73673	0.7094	0.77389	0.71795
6.0	0.73355	0.73504	0.76858	0.74359	0.76221	0.75214
7.0	0.77389	0.7094	0.74735	0.73504	0.77495	0.75214
8.0	0.76539	0.67521	0.76752	0.74359	0.77389	0.7265
9.0	0.76008	0.73504	0.77813	0.71795	0.77601	0.76923

As epoch increase k = 60% is best for validation and this pattern not continues in much higher epochs

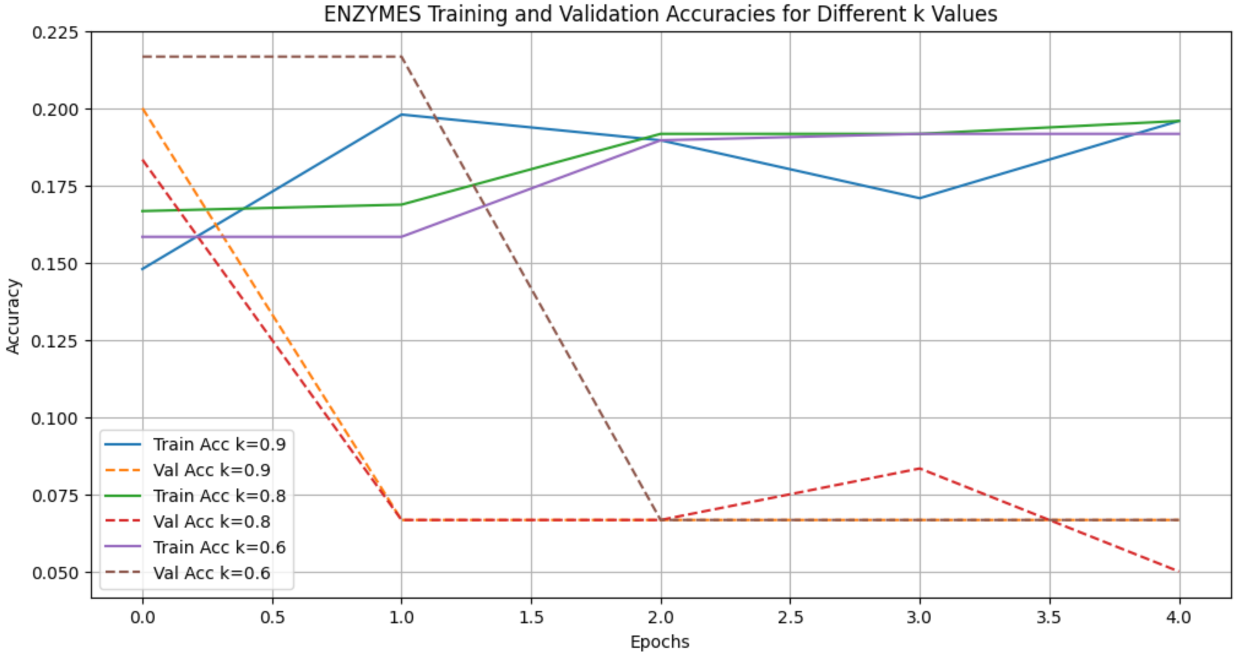
Analysis of ***ENZYMES*** - Binary Classification Task

Final accuracy after <u>1000</u> epoch with best_k 90% is : 67%

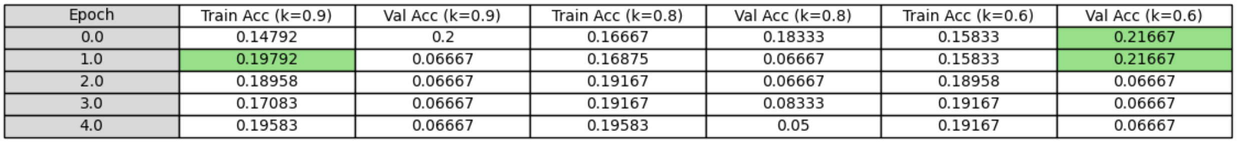
If number of epochs will be increased then it will improve the accuracy too.

We are going to analyze the loss and accuracy results on following points:
- cluster size in pooling layer i.e. m1 and m2 values
- By changing number of epochs (taking less number of epochs for visualization)

For epoch = 5 m1 = 6 and m2 = 3



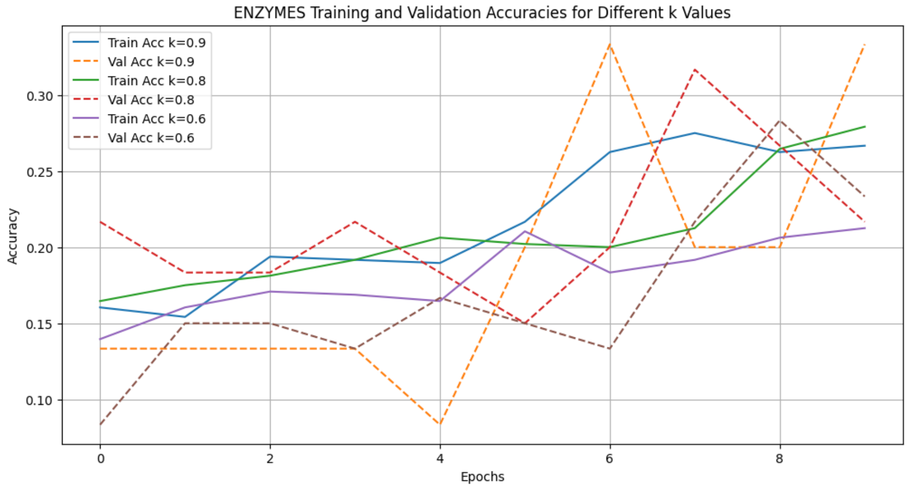
Training and Validation Accuracies

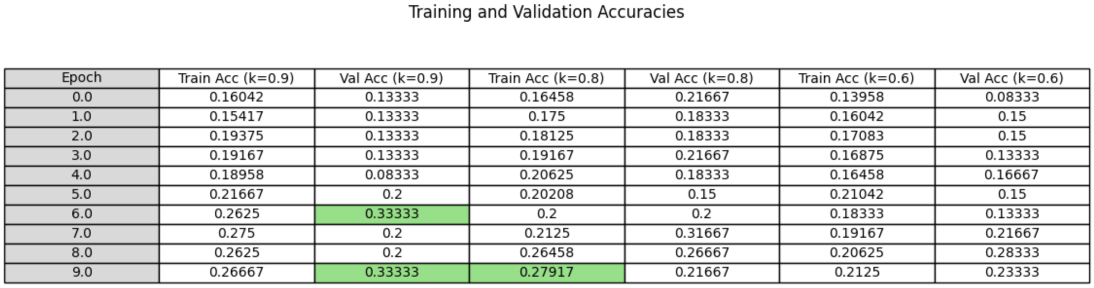


Epoch	Train Acc (k=0.9)	Val Acc (k=0.9)	Train Acc (k=0.8)	Val Acc (k=0.8)	Train Acc (k=0.6)	Val Acc (k=0.6)
0.0	0.14792	0.2	0.16667	0.18333	0.15833	0.21667
1.0	0.19792	0.06667	0.16875	0.06667	0.15833	0.21667
2.0	0.18958	0.06667	0.19167	0.06667	0.18958	0.06667
3.0	0.17083	0.06667	0.19167	0.08333	0.19167	0.06667
4.0	0.19583	0.06667	0.19583	0.05	0.19167	0.06667

This data table sis showing data for low number of epoch but from this we can understand that validation accuracy is very less compare to the train accuracy but after further analysis test accuracy is much better than validation accuracy. From data we can clearly see that validation accuracy is better for 60% k value in initial epochs and for train k value is 90% best.

For epoch = 10 m1 = 6 and m2 = 3



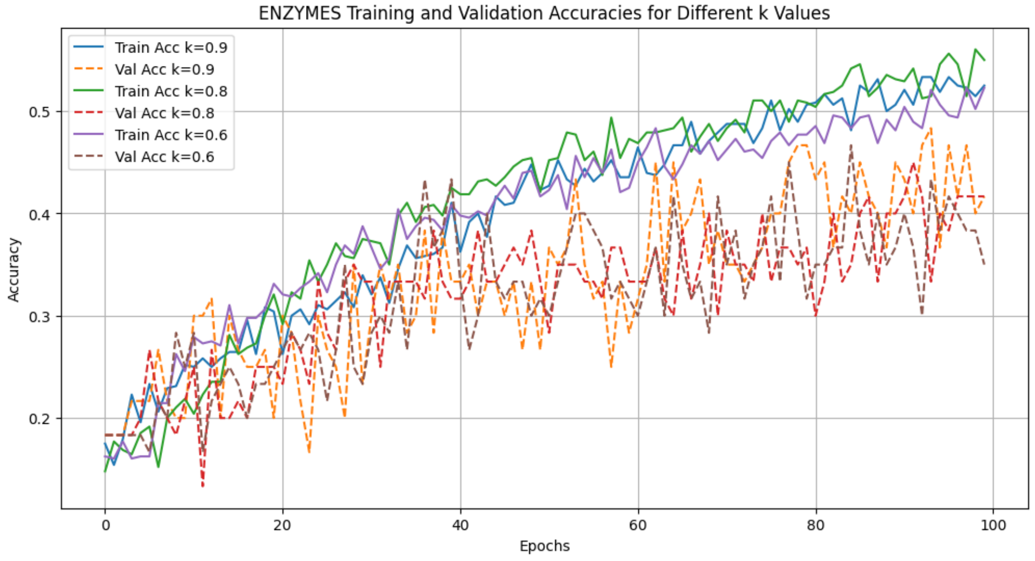


Training and Validation Accuracies

Epoch	Train Acc (k=0.9)	Val Acc (k=0.9)	Train Acc (k=0.8)	Val Acc (k=0.8)	Train Acc (k=0.6)	Val Acc (k=0.6)
0.0	0.16042	0.13333	0.16458	0.21667	0.13958	0.08333
1.0	0.15417	0.13333	0.175	0.18333	0.16042	0.15
2.0	0.19375	0.13333	0.18125	0.18333	0.17083	0.15
3.0	0.19167	0.13333	0.19167	0.21667	0.16875	0.13333
4.0	0.18958	0.08333	0.20625	0.18333	0.16458	0.16667
5.0	0.21667	0.2	0.20208	0.15	0.21042	0.15
6.0	0.2625	0.33333	0.2	0.2	0.18333	0.13333
7.0	0.275	0.2	0.2125	0.31667	0.19167	0.21667
8.0	0.2625	0.2	0.26458	0.26667	0.20625	0.28333
9.0	0.26667	0.33333	0.27917	0.21667	0.2125	0.23333

As epoch increase k = 90% is best for validation and this pattern continues in much higher epochs.

For 100 epoch :


ENZYMES Training and Validation Accuracies for Different k Values
Train Acc k=0.9
Val Acc k=0.9
Train Acc k=0.8
Val Acc k=0.8
Train Acc k=0.6
Val Acc k=0.6
Accuracy
Epochs