GRAPH CONVOLUTIONAL NETWORKS

CMU 11441/11741: MACHINE LEARNING WITH GRAPHS

Due date: 04/09/2024, 11:59 PM EST

https://github.com/cmu-ml4graph/gcn_assignment_s2024

https://www.overleaf.com/read/czfccfjtrczk#db25dc

Instructions

- **Allowed libraries:** This assignment involves implementing graph convolutional networks. You are **not** allowed to use any libraries that implement GCNs out of the box (like Pytorch-geometric). It is allowed to use autodiff libraries like Pytorch/Tensorflow. We highly recommend using Python + Pytorch for this assignment.
- **Getting feedback**: You can create a **private** fork of the repository on GitHub and add the TAs as collaborators (usernames: Edward-Sun). This might help you in asking questions without having to copy-paste your code on piazza (you can just reference Github code/copy permalink). You can use these instructions or just copy-paste the code into a new repository.
- **Posting your solutions online**: As with all the other assignments, please do not share your solutions publicly.
- **Statement of Assurance**

  1. Did you receive any help whatsoever from anyone in solving this assignment?
  2. Did you give any help whatsoever to anyone in solving this assignment?
  3. Did you find or come across code that implements any part of this assignment?

1 GCN Review (30 points)

Q1. What is the big-O time complexity of the computation expressed in Equation **??** in terms of $|\boldsymbol{V}|$, $|\boldsymbol{E}|$, d, k, and L? Your expression should not contain any other term. Assume $d < k$.

Solution

Q2. What is the space complexity of the computation expressed in Equation **??** in terms of $|\boldsymbol{V}|$, $|\boldsymbol{E}|$, d, k, and L (assume intermediate terms are saved)? Your expression should not contain any other term.

Solution

2 Graph Exploration (20 points)

Solution

Graph	Karate	Cora	Citeseer
Max in-degree	18	x	x
Min in-degree	2	x	x
Average in-degree	5.58	x	x
# nodes	34	x	x
# edges	190	x	x
Node feature dim	34	x	x

Table 1: Graph statistics

3 Node classification

3.1 Implementation (60 points)

Solution

Graph	Accuracy %	Loss
KARATE	100	0
CORA	x	x
CITESEER	x	x

Table 2: Node classification results

3.2 Varying L (20 points)

For both CORA and CITESEER, modify the **GNN** to include $L = 3, 4, 5, 6$ layers and plot the loss and accuracy vs. L. Summarize your observations in 2-3 lines.

Solution

3.3 Topological features vs. inbuilt features (20 points)

Solution

4 Link prediction

4.1 Training data for link prediction (20 points)

A. **Solution**

Graph	# Positive edges	# Negative edges
KARATE	190	190
CORA	x	x
CITESEER	x	x

Table 3: Training data statistic for link prediction

B. How is the training data for link prediction created? Please explain in 2-3 lines.

Solution

4.2 Implementation (80 points)

Solution

Graph	Accuracy %	Loss
KARATE	51.34	1.008
CORA	x	x
CITESEER	x	x

Table 4: Link Prediction Results

5 Graph classification

5.1 Graph Statistics (10 points)

Solution

Graph	MUTAG	ENZYMES
Num graphs	141	x
Avg. num nodes	18.85	x
Avg. num edges	94.04	x
Node feature dim	8	x

Table 5: Graph statistics for the graph classification datasets

5.2 Implementation (90 points)

Solution

Graph	MUTAG			ENZYMES		
	P	R	F1	P	R	F1
Mean-pooling	x	x		x	x	x
Max-pooling	84	83	83	x	x	x
Last-node pooling	x	x	x	x	x	x

Table 6: Graph classification results. Please use macro-averages to report the precision, recall, and F1 score for ENZYMES.

References