Weicong Feng

Professor Konstantinos Krampis

BIOL79303

11/24/2021

Assignment 10

Assignment 10

Due date: Wednesday, 11/25/2021

**Each of the questions below is worth 10 points.**

Simply write out each answer in 2-3 short sentences. If the question requires you to write math formulas, you can use Latex or write them by hand, but make sure you have clear writing and the pictures you get of the writing are also clear, and aligned to the rest of the document.

**Question 1.** How does the adjacency matrix represent the structure of a graph? How do adjacency matrices differ in directed versus un-directed graphs?

A picture containing watch

Description automatically generated

*Ans: Rows and columns represent nodes of the graphs in its adjacency matrix. While two nodes are connected, it shows non-zero value; otherwise, it could be zero. For example, as shown above, both X1,2 and X2,1 are 1 because 1 connects 2; while both X1,8 and X8,1 are zero because 1 and 8 are disconnected.*

*Adjacency matrix is symmetric along its main diagonal in undirected graphs, while directed graphs are not.*

**Question 2.** Do the graphs represent Euclidean data, and why do we have such an interest to being our data to Euclidean space? What would be the process or method for achieving this?

*Ans: Yes, the graphs can represent Euclidean data.*

**Question 3.** What are graph embeddings? Which are the overall components of the algorithms / procedures used for graph embedding. List a couple of algorithms that are used for graph embeddings.

*Ans: Graph embeddings are the transformation of property graphs to a low-dimensional vector or a set of vectors that can represent the graph structure information. It typically includes 4 components: similarity, similarity function, encoder, decoder. DeepWalk, node2vec, SDNE are common algorithms used for graph embeddings.*

**Question 4.** If we have a set of text-based data, how would we go about putting them in a data format that can used as input to neural networks? What kind of input would that be? List a couple of algorithms that are used for word embeddings.

*Ans: One hot encoded vector is typically used for presenting text-base data to input into a neural network. The common word embeddings algorithms include word2vec, GloVe, DeepWalk.*

**Question 5.** What is similarity in the embedding space of a graph means? How do you define similarity?

*Ans: The similarity means the nodes are close, connecting with edge, having same connected node, or around by similar network structure in the graph and is present as cluster closer in embedding space.*

**Question 6.** Define the encoder and the decoder in graph embedding. Can we have different types of encoders, and briefly what would be the differences between them?

*Ans: In graph embedding network, encoder translates the graph network to low-dimensional representation, such as vectors; while the decoder translates the embedded representation and tried to reconstruct and proximate to original graph network. Shallow encoder simply projects the graph to vectors, graph-regularization encoder exploits only the node features and labels, auto-encoding encoder component leverages only the graph structure, neighborhood aggregation encoder exploits both the node features and the graph structure in the propagation step.*

**Question 7.** What type of biological data are available that can be represented as graphs? Give a couple of examples in each case.

*Ans: The gene, disease, RNA and known association between them can be represented as graph.*

**Question 8.** Which type of predictions / learning tasks we can do with graph neural networks and biological data? Give a couple of examples in each case.

*Ans: Graph neural networks broadly applied to different biological data, for example, disease association prediction, RNA-Disease association, Disease-Gene association, image segmentation, and so on.*

**Question 9.** What is the issue with the structure of graphs that prevents the Graph Convolutional Networks (GCN), to be applied similarly to images in Convolutional Neural Networks (CNN)? Is graph embedding a GCN or is it a method that can be used to aid the GCN?

*Ans: The shallow depth prevents GCN from playing a more important role. In existing level, the GCN is difficult to be more than 3 levels. In general, more complex data and larger graph need more deeper neural network. Graph embedding is used to aid the GCN.*

**Question 10.** In the Kipf and Welling formula we discussed in class shown in one of the papers, how many layers we have in the neural network represented by the formula? How can we put in this formula in context with our previous matrix multiplications we were doing with our non-graph neural networks?

*Ans: There are two layers inside that formula. Most of this formular is same as the non-graph neural networks. The only difference is multiple an adjacency matrix.*