CSE8803/CX4803 - Homework 5 - Spring 2022

Submission guidline:

- Submit a `pdf` file that includes your answer of the theory questions.
- See the `ipynb` notebook for the submission of coding assignment.

1 Basic concepts of graphs [15 pts]

Consider the following undirected, unweighted graph in Figure 1 and answer the following questions. *Please show your work for all subproblems. Olny showing the numerical results will not receive points.*

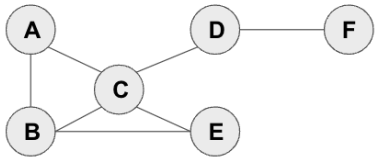


Figure 1: Input graph

(1) [5pts] In lecture, we introduced the betweenness centrality of a node v, which is defined as the sum of the fraction of all-pairs shortest paths that pass through v:

$$c_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{1}$$

where σ_{st} is the total number of shortest paths from node s to t and $\sigma_{st}(v)$ is the number of those paths that pass through v (not where v is an end point). Calculate the betweeness centrality for nodes C, D, and F in the input graph.

(2) [2pts] Recall that closeness centrality of a node v is the reciprocal of total shortest path distance between v and other nodes:

$$c_D(v) = \frac{1}{\sum_{u \neq v} d(u, v)} \tag{2}$$

where $d(u, v)$ is the distance between nodes u and v. Calculate the closeness centrality for nodes C and E in the input graph.

(3) [8pts] We introduced the concepts of graph isomorphism and graphlets in our lecture. Consider the graphlets with size up to 4 nodes (Figure 2). There are 15 unique graphlets in total (numbered from 0 to 14). Find which graphlets are contained by node C in our input graph. (Recall that we say node X in a graph has graphlet k if at least one induced subgraph covering node X of this graph is isomorphic with graphlet k.) For each unique graphlet that node C has, count its number of occurrence in the graph and list the nodes that form every occurrence (see the example below). You can use the index labeled in Figure 2 to refer to each graphlets.
For example: when you state that node A in Figure 1 contains graphlet 0, you also need to say that this graphlet occurs 2 times in the graph, which are A-B and A-C. (Only saying that node A has 2 subgraphs that match graphlet 0 but not listing what are them will not receive the full points.)

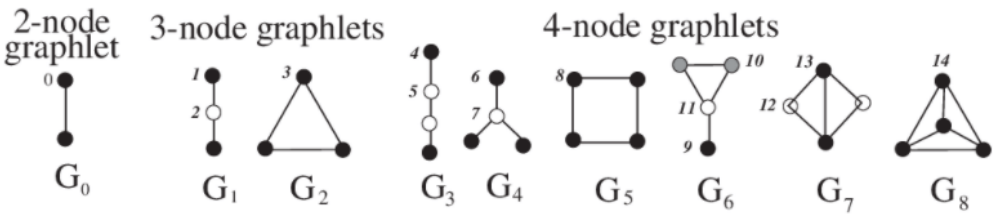


Figure 2: Graphlets with size up to 4 nodes.

Anwser:

(1) Node C (2 pts):

	A-D	A-E	A-F	B-D	B-F	D-E	E-F
#shotest path passing C	1	1	1	1	1	1	1
#shotest path	1	2	1	1	1	1	1

The betweenness centrality of node C is

$$\frac{1}{1}+\frac{1}{2}+\frac{1}{1}+\frac{1}{1}+\frac{1}{1}+\frac{1}{1}+\frac{1}{1}=6.5.$$

Node D (2 pts):

	A-F	B-F	C-F	E-F
#shotest path passing D	1	1	1	1
#shotest path	1	1	1	1

The betweenness centrality of node D is

$$\frac{1}{1}+\frac{1}{1}+\frac{1}{1}+\frac{1}{1}=4.$$

Node F (1 pt): F is not on the shortest path of any other pairs of nodes. So the betweenness centrality of node F is 0.

(2) Node C (1 pt): The shortest path distance of C to other nodes: A: 1, B: 1, D: 1, E: 1, F: 2. The closeness centrality of C is

$$\frac{1}{1+1+1+1+2}=\frac{1}{6}.$$

Node E (1 pt): The shortest path distance of C to other nodes: A: 2, B: 1, C:1, D: 2, F: 3. The closeness centrality of C is

$$\frac{1}{2+1+1+2+3}=\frac{1}{9}.$$

(3) Node C has 8 graphlets:

- Graphlet 0 (4 times): A-C, B-C, D-C, E-C;
- Graphlet 1 (once): C-D-F;
- Graphlet 2 (4 times): A-C-E, B-C-D, D-C-E, A-C-D;
- Graphlet 3 (2 times): A-B-C, B-C-E;
- Graphlet 5 (3 times): B-C-D-F, A-C-D-F, E-C-D-F;

- Graphlet 7 (once): A-C-D-E;
- Graphlet 11 (2 times): A-B-C-D, B-E-C-D;
- Graphlet 13 (once): A-B-C-E.

Rubric: to receive full points, you should list all graphlet types that C contains, count the occurrences of each graphlet correctly, and list the node sets for each occurrence correctly. -1 for each missing graphlet (up to 8 points); -1 for each wrong counts (up to 8 points); and -1 for each wrong node sets (up to 5 points).

2 Gradient descent [15 pts]

Logistic regression is one of the widely-used model for classification. The logistic regression model $h_{\boldsymbol{\theta}}(x)$ has the form

$$\hat{y} = h_{\boldsymbol{\theta}}(x) = \frac{1}{1 + e^{-\boldsymbol{\theta}^T \boldsymbol{x}}}, \tag{3}$$

where the vector $\boldsymbol{x}$ is the input, $\hat{y}$ is the output predicted by the model, and the vector $\boldsymbol{\theta}$ is the weights of the model. We simplify the setting such that there is only one training data point $(\boldsymbol{x}, y)$. Note that here $\boldsymbol{x} = [x_1, x_2, \ldots, x_m]^T$ is a m-dimensional features vector of the input and $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_m]^T$ is a vector of weights.

(1) [1pt] In class, we have seen the cross-entropy loss, a common loss function for classification problems. Write down this loss $L(y, \hat{y})$ in terms of y and $\hat{y}$.

(2) [4pts] The logistic regression is closely related to the sigmoid function, which is given by

$$g(z) = \frac{1}{1 + e^{-z}}. \tag{4}$$

It is easy to see that the logistic model $h_{\boldsymbol{\theta}}(x)$ can be expressed in terms of the sigmoid function: $h_{\boldsymbol{\theta}}(x) = g(\boldsymbol{\theta}^T \boldsymbol{x})$. When running the backpropagation algorithm to learn the weights of the logistic model, we need to compute the derivative of the simoid function $g'(z)$. Show that

$$g'(z) = g(z)(1 - g(z)) \tag{5}$$

(3) [3pts] Show that $1 - g(z) = g(-z)$.

(4) [6pts] Derive $\nabla_{\theta_j} L$, i.e., the gradient of the loss with respect to θ_j. (Hint: you may what to apply the chain rule and reuse the fact in Eq. 5).

(5) [1pt] After backpropagation, we can use the computed gradients to update every parameter θ_j ($j = 1, \ldots, m$) using gradient descent. Write down the update step of gradient descent for θ_j.

Answer:

(1) Cross-entropy (CE) loss:

$$L(y, \hat{y}) = -y \log \hat{y} - (1 - y) \log(1 - \hat{y}).$$

(-0.5 point if written as $y \log \hat{y} + (1 - y) \log(1 - \hat{y})$.)

(3) Using the chain rule

$$\begin{aligned} g'(z) &= \frac{d}{dz}g(z) = \frac{d}{dz}\left[\frac{1}{1+e^{-z}}\right] \\ &= \frac{d}{dz}(1+e^{-z})^{-1} \\ &= -1 \times (1+e^{-z})^{-2}(-e^{-z}) \\ &= \frac{e^{-z}}{(1+e^{-z})^2} \\ &= \frac{1}{1+e^{-z}} \cdot \frac{e^{-z}}{1+e^{-z}} \\ &= \frac{1}{1+e^{-z}}\left(1 - \frac{1}{1+e^{-z}}\right) \\ &= g(z)(1-g(z)) \end{aligned}$$

(3)

$$\begin{aligned} 1 - g(z) &= 1 - \frac{1}{1+e^{-z}} \\ &= \frac{e^{-z}}{1+e^{-z}} \\ &= \frac{e^{-z}}{1+e^{-z}} \cdot \frac{e^{z}}{e^{z}} \\ &= \frac{1}{1+e^{z}} = g(-z) \end{aligned}$$

(4) First, by chain rule, we have

$$\nabla_{\theta_j} L = \frac{\partial L}{\partial \theta_j} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial \theta_j}$$

Recall that $\hat{y} = g(\boldsymbol{\theta}^T \boldsymbol{x})$, then

$$\begin{aligned} \frac{\partial L}{\partial \theta_j} &= \frac{\partial}{\partial \theta_j}\left[-y \log \hat{y} - (1-y)\log(1-\hat{y})\right] && \text{(subproblem (1))} \\ &= \frac{\partial}{\partial \theta_j}\left[-y \log g(\boldsymbol{\theta}^T \boldsymbol{x})\right] + \frac{\partial}{\partial \theta_j}\left[-(1-y)\log(1-g(\boldsymbol{\theta}^T \boldsymbol{x}))\right] && \text{(derivative of sum of terms)} \\ &= \left[-\frac{y}{g(\boldsymbol{\theta}^T \boldsymbol{x})} - (-1)\cdot\frac{1-y}{1-g(\boldsymbol{\theta}^T \boldsymbol{x})}\right]\frac{\partial}{\partial \theta_j} g(\boldsymbol{\theta}^T \boldsymbol{x}) && \text{(derivative of } \log f(x)) \\ &= \left[-\frac{y}{g(\boldsymbol{\theta}^T \boldsymbol{x})} + \cdot\frac{1-y}{1-g(\boldsymbol{\theta}^T \boldsymbol{x})}\right] g(\boldsymbol{\theta}^T \boldsymbol{x})[1-g(\boldsymbol{\theta}^T \boldsymbol{x})]x_j && (g'(z) = g(z)(1-g(z))) \\ &= \left[-y + g(\boldsymbol{\theta}^T \boldsymbol{x})\right] x_j && \text{(cancelling terms)} \end{aligned}$$

(5)

$$\begin{aligned} \theta_j &= \theta_j - \eta \nabla_{\theta_j} L \\ &= \theta_j - \eta\left[-y + g(\boldsymbol{\theta}^T \boldsymbol{x})\right] x_j \end{aligned}$$

where η is the learning rate. (-0.5 point if ignored the learning rate.)

3 Bayesian network [10 pts]

(a) [3pts] Consider a gene regulatory network modeled with a Bayesian network (Figure 3). Please factorize the joint probability distribution $P(X1, X2, X3, X4)$ according to the network.

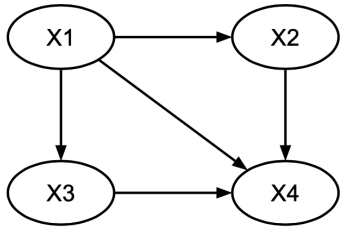


Figure 3: GRN

(b) [3pts] Now consider another gene regulatory network (Figure 4), we cannot factorize it directly due to the cycle structure. Dynamic Bayesian network (DBN) is a suitable model for such cases. Please reformulate the network into a DBN, and draw your DBN model below. You can assume that the network structure stays the same across different time points, so you only need to draw the nodes at time $t-1$ and t and the interactions.

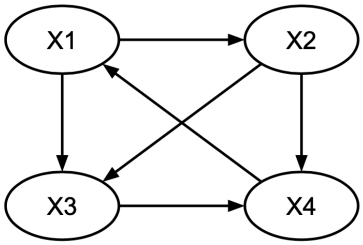


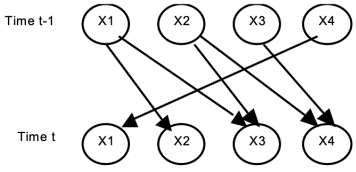
Figure 4: GRN

(c) [4pts] Consider a time series gene expression data of four genes $\mathcal{D} = \{\mathbf{X}[1], \mathbf{X}[2], \cdots, \mathbf{X}[T]\}$. And for each time point t, $\mathbf{X}[t] = [X1[t], X2[t], X3[t], X4[t]]$. Using the dynamic Bayesian network above (denoted as $\mathcal{G}$), factorize the likelihood function $P(\mathcal{D}|\mathcal{G})$. Note that you need to consider all time points in $P(\mathcal{D}|\mathcal{G})$.

Answer:

(a) [credit to: Bayesian Network Learning and Applications in Bioinformatics]

$$P(X1, X2, X3, X4) = P(X1)P(X2|X1)P(X3|X1)P(X4|X1, X2, X3) \tag{6}$$



(b)

(c)

$$
\begin{aligned}
& P(\mathcal{D}|\mathcal{G}) \\
=& \prod_{t=2}^{T} P(X1[t]|X4[t-1])P(X2[t]|X1[t-1])P(X3[t]|X1[t-1], X2[t-1])P(X4[t]|X2[t-1], X3[t-1])
\end{aligned} \tag{7}
$$

4 Programming assignment: Graph Neural Network [20 pts]

In this problem, you will implement a graph neural network and apply it on real-world data to predict properties of chemical molecules. Please see `hw5.ipynb` for instructions.