UIUC-CS512 "Data Mining Principles" (Fall 2022)
Second Midterm Exam

(Thursday, Nov. 17, 2022, 100 marks)

IMPORTANT Notes

- Please directly write down your solutions under every problem with brief explanations if necessary.
- The exam starts at 11:00 am and ends at 12:15 pm. Feel free to skip some 'hard' problems and distribute your time wisely.
- Please prepare your icard/ID and cheat-sheet when we collect your papers.
- Feel free to use the last blank page as scratch paper.

Name: NetID: Score:

1	2	3	4	Total

1 Back Propagation [25 points]

Consider a fully-connected neural network in Figure 1. For neuron i, we denote its input as I_i, output as O_i, and the weight for link with neuron j as w_{ij}. With input values I_1, I_2, I_3 and desired output T, answer the following questions using Sigmoid activation function ($O_j = \frac{1}{1+e^{-I_j}}$) and mean-square loss ($L = \frac{1}{2}(O - T)^2$). (You do not need to consider bias in this question)

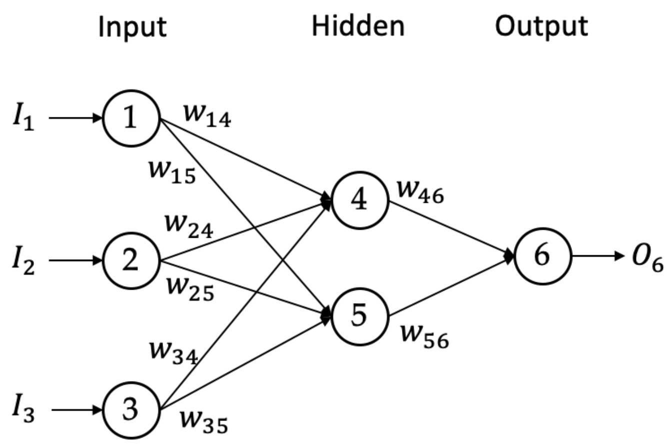

Figure 1: A fully-connected neural network

(a) (6 points) How to calculate I_4 in terms of I_1, I_2, I_3?

(b) (6 points) How to calculate the error $\delta_6 = \frac{\partial L}{\partial I_6}$ for the output neuron 6 in terms of O_6 and T? Show your intermediate steps.

(c) (8 points) How to calculate the error $\delta_4 = \frac{\partial L}{\partial I_4}$ for the hidden neuron 4 in terms of O_4 and δ_6? Show your intermediate steps.

(d) (5 points) If we apply dropout on non-output neurons in Figure 1 (i.e., neuron 1-5), how mainly possible distinct and connected networks are there? (connected network means at least one neuron is active in each layer)

2 Convolutional Neural Network [30 points]

(a) (5 points) Perform the convolution below with stride 1 and no zero-padding, i.e. calculate the image C.

$$
\begin{array}{|c|c|c|c|}
\hline
\mathbf{0} & 1 & 0 & 2 \\
\hline
1 & 0 & 1 & 0 \\
\hline
0 & 2 & 0 & 0 \\
\hline
\end{array}
\;*\;
\begin{array}{|c|c|}
\hline
0 & 1 \\
\hline
1 & \mathbf{1} \\
\hline
\end{array}
\;=\;
\begin{array}{|c|c|c|}
\hline
\mathbf{?} & ? & ? \\
\hline
? & ? & ? \\
\hline
\end{array}
$$

$$\text{image A} \quad * \quad \text{kernel B} \quad = \quad \text{image C}$$

(b) (10 points) A CNN consists of N complex layers. Each complex layer consists of a convolution with a 3×3 kernel (stride=2, no padding), a sigmoid activation, and a max pooling layer (stride = 2, kernels size = 2×2). The shape of the input image is 512×512. Compute the (spatial) size of the layer after the first and second pooling. When is the image of size 1×1, i.e. what is the maximum value for N? (throw the extra data during the convolution and pooling process)

(c) (5 points) How many paramenters do we have for the model in (b)?

(d) (10 points) Given a 1-d grid/sequence of input $[x_1, x_2, x_3, x_4, x_5]$ where x_i is a scalar. Three different models are used in Figure 2 which are MLP, CNN and RNN. The computation graph of these three models are shown in Figure 2. How many parameters do these three models have respectively? (No bias for CNN and MLP. The formula of RNN is shown in the Figure. h_i and s_i are scalars.)

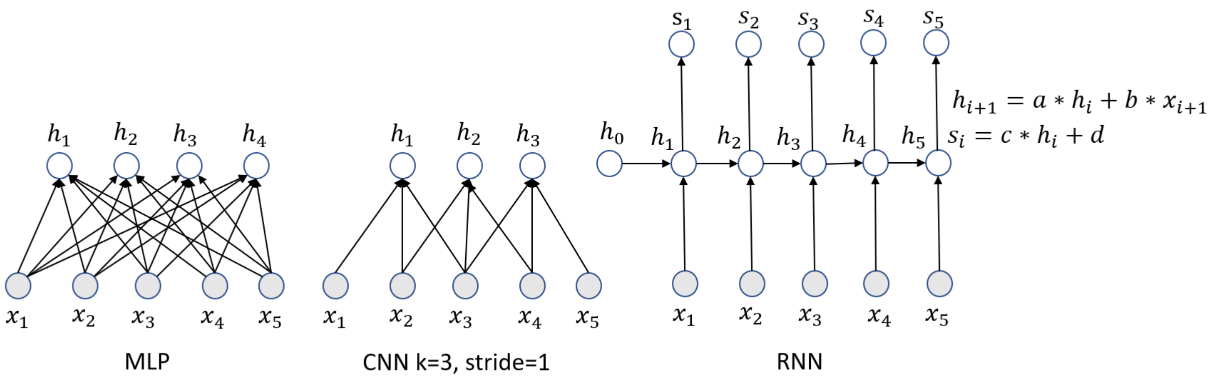

Figure 2: Compare the number of model parameters.

3 2-way Spectral Graph Partitioning [20 points]

An undirected unweighted graph (i.e., its adjacency matrix $\mathbf{A} = \mathbf{A}^T$) is given as Figure 3 shows. In the following questions, all the indices (of matrices and vectors) are consistent with the indices of nodes in the Figure 3.

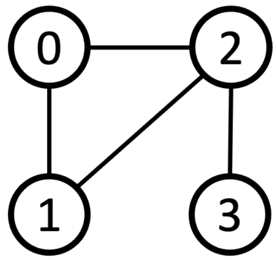

Figure 3: An undirected graph

(a) (7 points) Given a node membership vector $\mathbf{q}$ where $\mathbf{q}[i] = 1$ denotes node i is in cluster A and $\mathbf{q}[i] = -1$ denotes node i is in cluster B. If we define the CutSize as

$$\texttt{CutSize} = \frac{1}{4} \sum_{i,j} (\mathbf{q}[i] - \mathbf{q}[j])^2 \mathbf{A}[i,j]. \tag{3.1}$$

Please list all the 2-way graph partitioning cases (excluding the 'None vs. all nodes' cases) and their corresponding CutSizes.

(b) (4 points) What is the (unnormalized) graph Laplacian of the given graph?

(c) (4 points) Assume that the eigenvalues and eigenvectors of the graph Laplacian of the given graph are as Table 1 shows.

Table 1: Eigenvalues/vectors of the graph Laplacian

eigenvalues	**eigenvectors**
4	$[-0.29, -0.29, -0.87, -0.29]$
3	$[0.71, -0.71, 1.4 \times 10^{-15}, 5.1 \times 10^{-16}]$
0	$[-0.5, -0.5, -0.5, -0.5]$
1	$[-0.41, -0.41, -3.3 \times 10^{-16}, 0.82]$

What is the best 2-way partition of the minimal `CutSize` according to the spectral graph partitioning method (MinCut)?

(d) (5 points) We are given a larger graph as follows. If we apply the spectral graph partitioning method (MinCut), and use a certain eigenvector $\mathbf{v}$ of its graph Laplacian matrix to find the cut, what is the corresponding eigenvalue of the eigenvector $\mathbf{v}$?

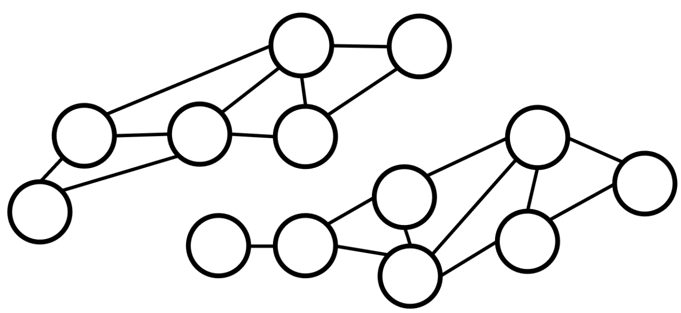

Figure 4: A larger undirected graph

4 Outlier Detection [25 points]

(a) (5 points) If sample o is defined as a $A(r, \pi)$-outlier if

$$\frac{||\{o'|o' \in D, o' \neq o, dist(o, o') \leq r\}||}{||D||} \leq \pi, \tag{4.2}$$

where $||\cdot||$ denotes the cardinality, D is the whole dataset, r is the distance threshold, and π is the fraction threshold. Is this outlier definition distance-based or density-based?

(b) (10 points) Given a 1-D dataset as: $\{-5, -4, 0, 3, 4\}$, report which nodes are the $A(2, 0.1)$-outlier and show all the intermediate results. Here we use the $dist(o, o') = |o - o'|$ as the distance function.

(c) (5 points) If the data distribute as Figure 5 shows in the feature space, does the outlier metric defined in (a) work? Briefly explain it.

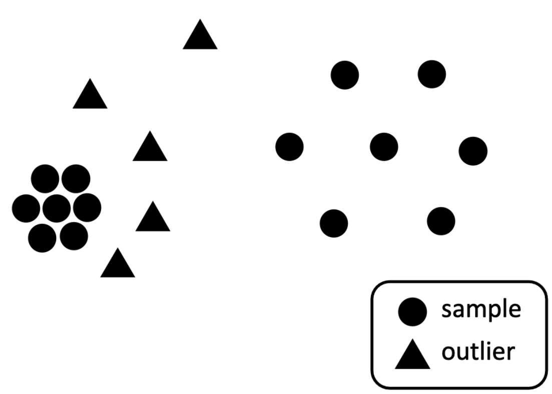

Figure 5: Data distribution.

(d) (5 points) Assume the data distribute as Figure 6 shows in the feature space. If we assume that the data is generated by a Gaussian distribution, which outlier(s) will be detected? If we assume that the data is generated by a mixture of two Gaussian distributions, which outlier(s) will be detected? Briefly explain it.

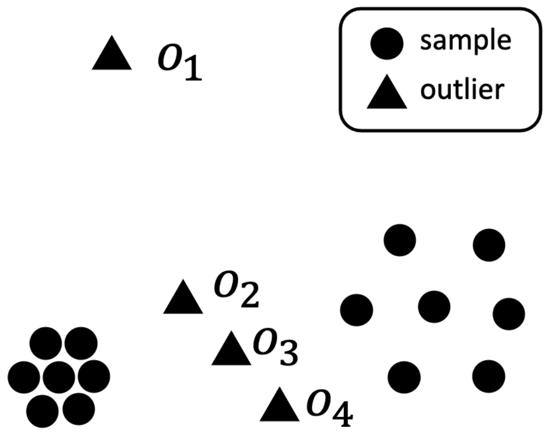

Figure 6: Data distribution.

Blank Scratch Paper