CS781 Endsem Exam (Autumn 2023)

Max marks: 50 **Duration: 180 mins**

- *The exam is open book and notes. However, you are not allowed to search on the internet or consult others over the internet for your answers.*
- *Be brief, complete and stick to what has been asked.*
- *Unless asked for explicitly, you may cite results/proofs covered in class without reproducing them.*
- ***If you need to make any assumptions, state them clearly.***
- ***Do not copy solutions from others. Penalty for offenders: FR grade.***

1. We have studied techniques for determining the robustness of deep neural networks w.r.t. perturbations of inputs, while the edge weights and biases remained fixed. In this question, we will explore robustness of networks w.r.t. perturbations of edge weights while the inputs and biases stay fixed.

   Consider the network shown in Fig. 1. All hidden and output layer nodes have bias 0. All hidden layer nodes use ReLU activation function. The edge weights shown are their nominal values, i.e. values without perturbation.

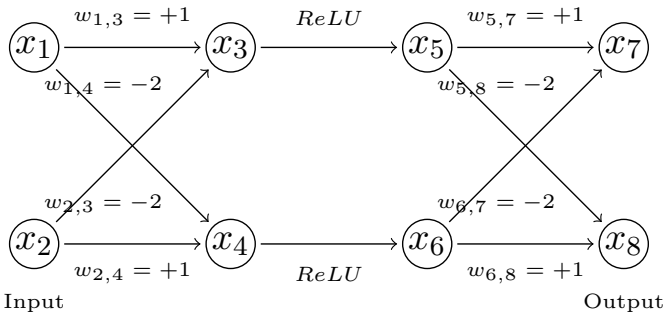

Figure 1: A neural network

   (a) *[7 marks]* Suppose the input to the network is $(x_1, x_2) = (0.5, 0.5)$. Without perturbing the inputs, suppose we perturb the edge weights w_{13} and w_{24} by adding a non-deterministic value chosen from $[-1, 1]$ to their nominal values. The perturbations of w_{13} and w_{24} are assumed to be independent of each other.

   Recall that in DeepPoly, we used a 4-tuple abstraction of the value at each node in the network. Using the same abstraction (i.e. linear upper bound expression, linear lower bound expression, concrete upper bound, concrete lower bound), find the 4-tuples for nodes x_7 and x_8. Note that you may have to include symbolic weights w_{13} and w_{24} as variables in the expressions for lower and upper bounds.

   (b) Suppose that in addition to w_{13} and w_{24}, the weight w_{57} is also perturbed by independently adding a non-deterministic values chosen from $[-1, 1]$ to its nominal value.

      i. *[10 marks]* We want to use a 4-tuple abstraction that is similar to what is used in DeepPoly, except that we will allow quadratic expressions for the upper and lower bounds, if needed. Find symbolic expressions in terms of w_{13}, w_{24} and w_{57} for upper and lower bounds of the value of x_7.

ii. *[3 marks]* Find as best concrete upper and lower bounds of the value of x_7 as you can from the expressions obtained above.

2. In class, we have studied how optimal binary decision trees for a given set of data points can be constructed by solving a system of constraints. In this question, we will take this endeavour a bit further and try to construct optimal *binary decision diagrams*, which are like binary decision trees (i.e. every non-leaf node is labeled by a feature, and has a left child and a right child), except that the final model can be a rooted DAG and does not necessarily have to be a tree. Figure 2 shows an example binary decision tree and a binary decision diagram for the same data set (shown in sub-question (b)).

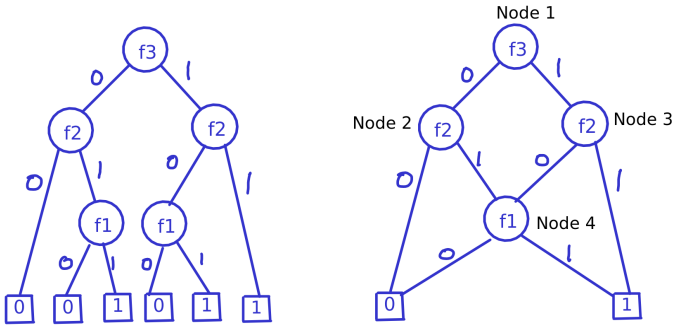

Figure 2: Decision tree and decision diagram

(a) *[10 marks]* Show that for every $n > 0$, there exists a data set over n binary features $f_1, f_2, \ldots f_n$ and a single binary decision, such that the smallest binary decision tree for the data set has $\Omega(2^n)$ decision (or internal nodes), while the smallest binary decision diagram for the same data has $O(n)$ decision nodes.

This shows the advantage of binary decision diagrams over binary decision trees.

(b) Consider the data shown in the following table, where f_1, f_2, f_3 are binary features and d is a binary decision.

Data point	f_1	f_2	f_3	d
1	0	0	1	0
2	0	1	0	0
3	1	0	1	1
4	1	1	0	1
5	1	0	0	0
6	0	1	1	1

We wish to find an optimal decision diagram with 4 decision nodes (and two leaves labeled 0 and 1) for this data set by solving a system of constraints. As in the case of decision trees, we assume that

- Nodes of the decision diagram are numbered from 1 through 4, with 1 being the root.
- All non-leaf children of a node are numbered higher than the node itself.
- For every non-leaf node, if the feature labeling the node has value 0 for a given data point, then we move to the left child of the node en-route to a leaf to find the decision for the data point. Otherwise, if the feature value is 1, we move to the right child of the corresponding node.
- A node i is *truthful* for data point j iff starting from node i and reading the values of features in data point j, we can reach a leaf node labeled by the decision corresponding to data point j.

We use the following variables to construct a system of constraints such that finding a satisfying assignment of the constraints yields a decision diagram for the given data set.

Variable	True iff ...
$a_{i,j}$	Node i is labeled by feature f_j, $1 \leq i \leq 4$, $1 \leq j \leq 3$
$l_{i,j}$	Node j is the left child of node i, $1 \leq i \leq 3$, $i < j \leq 4$
$r_{i,j}$	Node j is the right child of node i, $1 \leq i \leq 3$, $i < j \leq 4$
$l0_i$	Leaf labeled 0 is the left child of node i, $1 \leq i \leq 4$
$l1_i$	Leaf labeled 1 is the left child of node i, $1 \leq i \leq 4$
$r0_i$	Leaf labeled 0 is the right child of node i, $1 \leq i \leq 4$
$r1_i$	Leaf labeled 1 is the right child of node i, $1 \leq i \leq 4$
$t_{i,j}$	Node i is truthful for data point j, $1 \leq i \leq 4$, $1 \leq j \leq 6$

(a) *[5 marks]* For the decision diagram shown in Fig. 2, list the values of all variables described above.

(b) *[2.5 × 3 marks]* Write propositional formulas in terms of the above variables to encode each of the following constraints:

i. Every non-leaf node i has two distinct children.

ii. Every non-root and non-leaf node i has at least one parent.

iii. For every non-leaf node i and for every data point j, the node is truthful for data point j iff the child of i reached on the value of feature labeling node i is also truthful for data point j.

(c) *[12.5 marks]* Argue why a satisfying assignment for the conjunction of constraints obtained above need not necessarily give an optimal (i.e. with minimum number of decision nodes) binary decision diagram for a given arbitrary data set (not necessarily the one given in this question)? Given reasons in support of your answer. Answers without reasons will fetch no marks.

(d) *[5 marks]* What additional constraints would you add to those obtained above to ensure that a satisfying assignment necessarily gives an optimal binary decision diagram for a given data set. Given reasons in support of your answer. Answers without reasons will fetch no marks.