# CS620 Quiz 3 (Spring 2021)

**Max marks: 35** **Due: Sun Apr 18, 7.30pm**

- *Be brief, complete and stick to what has been asked.*
- *Untidy presentation of answers, and random ramblings will be penalized by negative marks.*
- *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.***
- **Expected time to solve:** 1 **hr** 30 **mins.**

1. In the paper on DeepPoly, we have seen that values of individual variables are bounded above and below by linear inequalities involving other variables. Suppose we now restrict the kind of linear inequalities that can be used. Specifically, suppose we require that the coefficients of variables in every linear inequality must be an element of the set $\{0, 1, -1\}$. So you can't have an inequality like $x_8 \leq 0.5 \cdot x_3 - 1$ (because the coefficient of $x_3$ is neither 0, 1 nor $-1$), but you can have something like $x_8 \leq x_3 - x_5 + x_6 - 27$. Note that while the coefficients of all variables are in $\{0, 1, -1\}$ in the last inequality, the constant term $-27$ is not restricted. We call inequalities of this form *simple linear constraints.*

   (a) *[10 marks]* Suppose you are told that $L \leq x \leq U$, where $L$ and $U$ are constants. Give the best over-approximation for $y = \mathsf{ReLU}(x)$ you can come up with using simple linear constraints, when $x$ lies in $[L, U]$.

   Give justification why your over-approximation is the best possible, given the requirement of using simple linear constraints. Your answer must consider all possibilities of $L$ and $U$, such that $L \leq U$.

   (b) *[10 marks]* Let $x_\ell = \sum_{i \in Preds(\ell)} W_{i,\ell} \cdot x_i$ give the value of a node $x_\ell$ in a deep neural network in terms of the values of nodes $x_i$ for $i \in Preds(\ell)$, where $Preds(\ell)$ gives the indices of all nodes that have an edge to $x_\ell$, and $W_{i,\ell}$ is the weight of the neural network edge from node $x_i$ to node $x_\ell$. Assume that $i < \ell$ for all $i \in Preds(\ell)$.

   For every variable $x_j$, $(1 \leq j < \ell)$, suppose $\sum_{i=1}^{j-1} l_{i,j} \cdot x_i + L_j \leq x_j \leq \sum_{i=1}^{j-1} u_{i,j} \cdot x_i + U_j$ are simple linear constraints bounding the value of $x_j$, where all $l_{i.j}, u_{i,j}$ are elements of $\{0, 1, -1\}$, and $L_j, U_j$ are unrestricted constants.

   Given as good lower and upper bounds of $x_\ell$ as you can using simple linear constraints. Thus, you must give a simple linear constraint to serve as a lower bound for $x_\ell$ and a simple linear constraint to serve as an upper bound for $x_\ell$.

   Explain clearly why your simple linear constraints are sound lower and upper bounds, i.e. the value of $x_\ell$ can never violate these bounds.

2. Consider the simple neural network shown in Fig. 1 below. Assume that each node in the hidden and output layers uses a ReLU activation function. We will use the terminology used in the paper "On the Effectiveness of Interval Bound Propagation for Training Verifiably Robust Models" by Sven Gowal et al, that we have studied in class.

   (a) *[10 marks]* Suppose $y_{true}$ is the component of the output vector that has the highest value. For example, if the inputs $in_1$ and $in_2$ have the value 1 each, the outputs have values $out_1 = 3$ and $out_2 = 6$, and hence $y_{true} = 2$ in this case.

   For the example considered above, we wish to check if $y_{true}$ stays at 2 if $in_2$ has a value in $[0, 1]$ (instead of being 1). Use the interval bound propagation technique discussed in the paper mentioned above to determine the $\widehat{z}_K$ vector referred to in equation (10), page 4 of the above paper. For our purposes, $K = 2$ and $z_2 = (out_1, out_2)$.

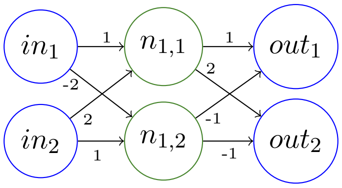

Figure 1: A simple DNN

(b) *[5 marks]* Referring to the loss function in equation 12, page 4 of the above paper, assume a simple loss function $\ell(z_K, y)$ that evaluates to 0 if the $y^{th}$ component of $z_K$ is the largest among all components of $z_k$, and evaluates to 1 otherwise.

For the example considered above and for the perturbation of $in_2$ considered above in part (a), what is the largest value of the loss function $L$ that can be obtained if we use values of $\kappa \in [0, 1]$. Does this answer change with the value of $\kappa$ for this example?