CS781 Midsem Exam (Autumn 2023)

Max marks: 45 **Duration: 120 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.***

Consider the neural network shown in Fig. 1. All hidden and output layer nodes have bias 0. All hidden layer nodes use ReLU activation function. For each hidden layer node x_i, we will use x_i^{pre} to denote the pre-activation value of x_i and $x_i^{post} = \texttt{R}eLU(x_i^{pre})$ to denote the post-activation value of x_i. Nodes x_5 and x_6 of the network give the scores of two possible output labels for a given input (x_1, x_2). Node x_7 is used to find the difference $x_5 - x_6$, so that we can figure out whether $x_5 \geq x_6$ or not by inspecting the sign of x_7.

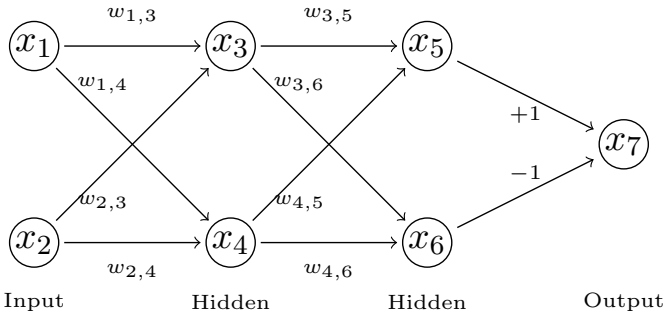

Figure 1: A neural network

1. Assume $w_{1,3} = w_{2,4} = w_{3,6} = w_{4,5} = 4$ and $w_{1,4} = w_{2,3} = w_{3,5} = w_{4,6} = -3$. Please draw the neural network carefully with the above weights annotated on the edges before proceeding further.

   (a) *[2 marks]* For $(x_1, x_2) = (0, 1)$, find the sign (+ve, 0 or -ve) of x_7.

   (b) *[3 marks]* Now consider perturbation of the above input, such that (x_1, x_2) lies in $\mathbb{B}_2\big((0,1), 1\big)$, i.e. in an ℓ_2-norm ball of radius 1 around $(0, 1)$. Using Holder's inequality, find lower and upper bounds of the pre-activation values of x_3 and x_4.

   (c) *[2 marks]* Using interval propagation and the pre-activation bounds of x_3 and x_4 obtained above, compute lower and upper bounds of the pre-activation values of x_5 and x_6.

   (d) *[10 marks]* Using the DeepPoly algorithm studied in class, find lower and upper bound linear expressions for x_7 in terms of x_1 and x_2. You must show each step of application of your algorithm clearly.

   (e) *[3 marks]* Find the tightest upper and lower numerical bounds of x_7 from the above expressions, and with knowledge of the fact that (x_1, x_2) lies in $\mathbb{B}_2\big((0,1), 1\big)$.

2. Now consider the same neural network with the following edge weights: $w_{1,3} = w_{2,4} = w_{3,6} = w_{4,5} = 2$ and $w_{1,4} = w_{2,3} = w_{3,5} = w_{4,6} = -1$. Please draw the neural network carefully with the above weights annotated on the edges before proceeding further. For this question, the inputs are perturbed such that $x_1 \in [-1, 1]$, $x_2 \in [-1, 1]$, i.e. $(x_1, x_2) \in \mathbb{B}_\infty\big((0, 0), 1\big)$.

   (a) *[10 marks]* We wish to use the α-CROWN approach to find bounds on x_7. Assuming that the slopes of the lower bounding approximations for each of the ReLUs in x_5 and x_6 are 1, and that for each of the ReLUs in x_3 and x_4 are α (the same parameter for both ReLUs), where $0 \leq \alpha \leq 1$, find lower and upper bounding linear expressions of x_7 (these can be in terms of α).

   Use simple interval propagation to obtain initial bounds on pre-activation values of hidden layer neurons.

   (b) *[3 marks]* Find the best numeric lower and upper bounds of x_7 that you can obtain by optimizing the above expressions over the allowed range of α.

   (c) *[10 marks]* Now suppose we fix $\alpha = 1$ in the above sub-problem, but wish to split the unstable neuron x_3 as is done in the β-CROWN approach. Therefore, we will use a β parameter for this neuron, where $\beta \geq 0$. All other unstable neurons are assumed to be left unsplit. For the same perturbation as above, find lower and upper bounding expressions of x_7 in terms of β. Solve the problem for both branches of the split of neuron x_3.

   As before, use simple interval propagation to obtain initial bounds on pre-activation values of hidden layer neurons.

   (d) *[2 marks]* Find the best numeric lower and upper bounds of x_7 that you can obtain by optimizing the above expressions over the allowed range of β.