Technical Report

No Author Given

No Institute Given

1 Proof

1.1 Proof of Theorem 1

Proof. Let $\Delta\mathbf{x} = \tilde{\mathbf{x}}^i_j - \mathbf{x}^i_j = \text{clamp}(\tilde{\mathbf{y}}^i_j, 0, t) - \text{ReLU}(\mathbf{y}^i_j)$ be the difference of two neurons in QNN and DNN after applying the activation function. Note that, here we use $\tilde{\mathbf{y}}^i_j$ (resp. $\mathbf{y}^i_j$) to denote the value of neuron before an activation function in QNN (resp. DNN). Let $\Delta = \tilde{\mathbf{y}}^i_j - \mathbf{y}^i_j$.

First, considering **Case 1** ($\text{UB}(S^{in}(\mathbf{x}^i_j)) \leq 0$), the neuron of DNN is always deactivated as 0. Hence, the output difference $\delta_{i,j} = S(\tilde{\mathbf{x}}^i_j)$.

Next, we consider following cases when the neuron in DNN is always activated, i.e., $\text{LB}(S^{in}(\mathbf{x}^i_j)) > 0$:

- **Case 2-1&2-2** ($\text{UB}(S^{in}(\tilde{\mathbf{x}}^i_j)) \leq 0$ or $\text{LB}(S^{in}(\tilde{\mathbf{x}}^i_j)) \geq t$): Since the neuron in QNN is always deactivated or clampped to t, we can get $\delta_{i,j} = -S^{in}(\mathbf{x}^i_j)$ or $\delta_{i,j} = t - S^{in}(\mathbf{x}^i_j)$.
- **Case 2-3** ($\text{LB}(S^{in}(\tilde{\mathbf{x}}^i_j)) \geq 0$ and $\text{UB}(S^{in}(\tilde{\mathbf{x}}^i_j)) \leq t$): Both of two neurons in QNN and DNN are activated and without clampped. Therefore, we have $\delta_{i,j} = \delta^{in}_{i,j}$.
- **Case 2-4** ($\text{LB}(S^{in}(\tilde{\mathbf{x}}^i_j)) < 0$ and $0 \leq \text{UB}(S^{in}(\tilde{\mathbf{x}}^i_j)) \leq t$): Since $\tilde{\mathbf{y}}^i_j$ is always smaller than t, we have $\Delta\mathbf{x} = \max(\tilde{\mathbf{y}}^i_j, 0) - \mathbf{y}^i_j = \max(\tilde{\mathbf{y}}^i_j - \mathbf{y}^i_j, -\mathbf{y}^i_j)$. Then, $\delta = \max(\delta^{in}_{i,j}, -S^{in}(\mathbf{x}^i_j))$.
- **Case 2-5** ($0 \leq \text{LB}(S^{in}(\tilde{\mathbf{x}}^i_j)) < t$ and $\text{UB}(S^{in}(\tilde{\mathbf{x}}^i_j)) > t$): Since $\tilde{\mathbf{y}}^i_j$ is always larger than 0, we have $\Delta\mathbf{x} = \min(\tilde{\mathbf{y}}^i_j, t) - \mathbf{y}^i_j = \min(\tilde{\mathbf{y}}^i_j - \mathbf{y}^i_j, t - \mathbf{y}^i_j)$. Hence, we have $\delta_{i,j} = \min(\delta^{in}_{i,j}, t - S^{in}(\mathbf{x}^i_j))$.
- **Case 2-6** (Otherwise): By case 2-4 & 2-5, we have $\Delta\mathbf{x} = \max(\min(\tilde{\mathbf{y}}^i_j, t), 0) - \mathbf{y}^i_j = \max(\min(\Delta, t - \mathbf{y}^i_j), -\mathbf{y}^i_j)$. Hence, we have $\delta_{i,j} = \max(\min(\delta^{in}_{i,j}, t - S^{in}(\mathbf{x}^i_j)), -S^{in}(\mathbf{x}^i_j))$.

Finally, we consider following cases when the neuron in DNN can be either activated or deactivated:

- **Case 3-1&3-2** ($\text{UB}(S^{in}(\tilde{\mathbf{x}}^i_j)) \leq 0$ or $\text{LB}(S^{in}(\tilde{\mathbf{x}}^i_j)) \geq t$): Similar to above, we have $\delta_{i,j} = -S(\mathbf{x}^i_j)$ or $t - S(\mathbf{x}^i_j)$ directly.
- **Case 3-3** ($\text{LB}(S^{in}(\tilde{\mathbf{x}}^i_j)) \geq 0$ and $\text{UB}(S^{in}(\tilde{\mathbf{x}}^i_j)) \leq t$): $\Delta\mathbf{x}^i_j = \tilde{\mathbf{y}}^i_j - \max(\mathbf{y}^i_j, 0) = \tilde{\mathbf{y}}^i_j + \min(-\mathbf{y}^i_j, 0) = \min(\tilde{\mathbf{y}}^i_j, \Delta)$. Then, we have $\delta_{i,j} = \min(S^{in}(\tilde{\mathbf{x}}^i_j), \delta^{in}_{i,j})$.
- **Case 3-4** ($\text{LB}(S^{in}(\tilde{\mathbf{x}}^i_j)) < 0$ and $0 \leq \text{UB}(S^{in}(\tilde{\mathbf{x}}^i_j)) \leq t$): We rewrite $\Delta\mathbf{x}$ as $\max(\tilde{\mathbf{y}}^i_j) - \max(\mathbf{y}^i_j)$:

- If $\Delta \leq 0$, $\Delta\mathbf{x} = \max(\mathbf{y}_j^i + \Delta) - \max(\mathbf{y}_j^i) \leq 0$. Then, we have $\Delta\mathbf{x} = 0$ when $\mathbf{y}_j^i \leq 0$, and $\Delta\mathbf{x} = \max(\mathbf{y}_j^i + \Delta, 0) - \mathbf{y}_j^i = \max(\Delta, -\mathbf{y}_j^i) \leq 0$ when $\mathbf{y}_j^i \geq 0$. Therefore, we have $\max(\mathrm{LB}(\delta_{i,j}^{in}), -\mathrm{UB}(S^{in}(\mathbf{x}_j^i))) \leq \Delta\mathbf{x} \leq 0$
- If $\Delta \geq 0$, $\Delta\mathbf{x} = \max(\tilde{\mathbf{y}}_j^i) - \max(\tilde{\mathbf{y}}_j^i - \Delta) \geq 0$. Then, we have $\Delta\mathbf{x} = 0$ when $\tilde{\mathbf{y}}_j^i \leq 0$, and $\Delta\mathbf{x} = \tilde{\mathbf{y}}_j^i - \max(\tilde{\mathbf{y}}_j^i - \Delta, 0) = \tilde{\mathbf{y}}_j^i + \min(\Delta - \tilde{\mathbf{y}}_j^i, 0) = \min(\Delta, \tilde{\mathbf{y}}_j^i) \geq 0$ when $\mathbf{y}_j^i \geq 0$. Therefore, we have $0 \leq \Delta\mathbf{x} \leq \min(\mathrm{UB}(\delta_{i,j}^{in}), \mathrm{UB}(S^{in}(\tilde{\mathbf{x}}_j^i)))$

Therefore, we have $\mathrm{LB}(\delta_{i,j}) = 0$ if $\mathrm{LB}(\delta_{i,j}^{in}) \geq 0$, and $\max(\mathrm{LB}(\delta_{i,j}^{in}), -\mathrm{UB}(S^{in}(\mathbf{x}_j^i)))$ otherwise. $\mathrm{UB}(\delta_{i,j}) = 0$ if $\mathrm{UB}(\delta_{i,j}^{in}) \leq 0$, and $\min(\mathrm{UB}(\delta_{i,j}^{in}), \mathrm{UB}(S^{in}(\tilde{\mathbf{x}}_j^i)))$ otherwise.

– **Case 3-5** ($0 \leq \mathrm{LB}(S^{in}(\tilde{\mathbf{x}}_j^i)) \leq t$ and $\mathrm{UB}(S^{in}(\tilde{\mathbf{x}}_j^i)) > t$): We rewrite $\Delta\mathbf{x}$ as $\min(\tilde{\mathbf{y}}_j^i, t) - \max(\mathbf{y}_j^i, 0)$. Then, $\Delta\mathbf{x} = t - \max(\mathbf{y}_j^i, 0) = \min(t - \mathbf{y}_j^i, t)$ when $\tilde{\mathbf{y}}_j^i \geq t$, and $\Delta\mathbf{x} = \tilde{\mathbf{y}}_j^i - \max(\tilde{\mathbf{y}}_j^i - \Delta, 0) = \tilde{\mathbf{y}}_j^i + \min(\Delta - \tilde{\mathbf{y}}_j^i, 0) = \min(\Delta, \tilde{\mathbf{y}}_j^i)$ when $\tilde{\mathbf{y}}_j^i \leq t$. Specifically, when $\tilde{\mathbf{y}}_j^i \geq t$:
  - If $\Delta \leq t$, then $\mathbf{y}_j^i = \tilde{\mathbf{y}}_j^i - \Delta \geq 0$, and we will have $\Delta\mathbf{x} = t - \mathbf{y}_j^i = t - \tilde{\mathbf{y}}_j^i + \Delta \leq \Delta \leq t$;
  - If $\Delta \geq t$, then $\Delta\mathbf{x} = t - \max(\mathbf{y}_j^i, 0) \leq t \leq \Delta$.

  Then, $t - \mathrm{UB}(S^{in}(\mathbf{x}_j^i)) \leq \Delta\mathbf{x} \leq \{\Delta, t\}$ for $\tilde{\mathbf{y}}_j^i \leq t$, and $\min(\mathrm{LB}(\delta_{i,j}^{in}), \mathrm{LB}(S^{in}(\tilde{\mathbf{x}}_j^i))) \leq \Delta\mathbf{x} \leq \{\Delta, t\}$ for $\tilde{\mathbf{y}}_j^i \leq t$. Finally, $\mathrm{LB}(\delta_{i,j}) = \min\big(\mathrm{LB}(\delta_{i,j}^{in}), \mathrm{LB}(S^{in}(\tilde{\mathbf{x}}_j^i)), t - \mathrm{UB}(S^{in}(\mathbf{x}_j^i))\big)$ and $\mathrm{UB}(\delta_{i,j}) = \min\big(\mathrm{UB}(\delta_{i,j}^{in}), t\big)$.

– **Case 3-6** (Otherwise):
  - If $\tilde{\mathbf{y}}_j^i < 0$, $\mathrm{UB}(\delta_{i,j}) = 0$ by case 3-1.
  - If $\tilde{\mathbf{y}}_j^i \geq 0$, $\mathrm{UB}(\delta_{i,j}) = \min\big(\mathrm{UB}(\delta_{i,j}^{in}), t\big)$ by case 3-5.
  - If $\tilde{\mathbf{y}}_j^i \geq t$, $\mathrm{LB}(\delta_{i,j}) = t - \mathrm{UB}(S^{in}(\mathbf{x}_j^i))$ by case 3-2;
  - If $\tilde{\mathbf{y}}_j^i \leq t$, $\mathrm{LB}(\delta_{i,j}) = 0$ if $\mathrm{LB}(\delta_{i,j}^{in}) \geq 0$, and $\max(\mathrm{LB}(\delta_{i,j}^{in}), -\mathrm{UB}(S^{in}(\mathbf{x}_j^i)))$ otherwise by case 3-4.

  Then, we get the lower bound as

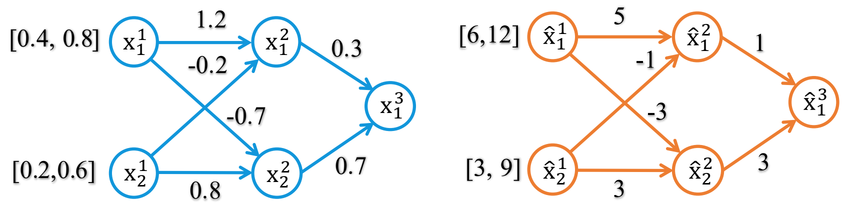
$$\mathrm{LB}(\delta_{i,j}) = \min\big(t - \mathrm{UB}(S^{in}(\mathbf{x}_j^i)), 0, \max(\mathrm{LB}(\delta_{i,j}^{in}), -\mathrm{UB}(S^{in}(\mathbf{x}_j^i)))\big),$$

  and upper bound as

$$\mathrm{UB}(\delta_{i,j}) = \max\big(\min(\mathrm{UB}(\delta_{i,j}^{in}), t), 0\big) = \mathrm{clamp}(\mathrm{UB}(\delta_{i,j}^{in}), 0, t).$$

2 Examples

Now we give the illustration examples for Section 3 and Section 4, both of which are based on DNN $\mathcal{N}_e$ and QNN $\widehat{\mathcal{N}}_e$ given in Fig. 1. Note that, the weights of DNN $\mathcal{N}_e$ and QNN $\widehat{\mathcal{N}}_e$ used here is little different from that in the paper (i.e., Fig. 1) for simplification consideration. The input region for all the examples in this section for QNN $\widehat{\mathcal{N}}_e$ is $R((9,6),3) = \{(x,y) \in \mathbb{Z}^2 \mid 6 \leq x \leq 12, 3 \leq y \leq 9\}$.



(a) DNN $\mathcal{N}_e$. (b) QNN $\widehat{\mathcal{N}}_e$.

Fig. 1. A 3-layer DNN $\mathcal{N}_e$ and its quantized version $\widehat{\mathcal{N}}_e$.

2.1 Example for Section 3

Example 1. Consider the DNN $\mathcal{N}_e$ and QNN $\widehat{\mathcal{N}}_e$ given in Fig. 1(a) and Fig. 1(b), where $2^{-F_{in}} = 1/16$, $\mathcal{C}_h^{\rm ub} = 1/15$. We set the input region for QNN $\widehat{\mathcal{N}}_e$ as $R((9,6),3) = \{(x,y) \in \mathbb{Z}^2 \mid 6 \le x \le 12, 3 \le y \le 9\}$.

By symbolic interval analysis, we have $S(\mathbf{x}_1^1) = S^{\rm in}(\mathbf{x}_1^1) = [0.4, 0.8]$, $S(\mathbf{x}_2^1) = S^{\rm in}(\mathbf{x}_2^1) = [0.2, 0.6]$, $S(\mathbf{x}_1^2) = S^{\rm in}(\mathbf{x}_1^2) = [0.36, 0.92]$, $S(\mathbf{x}_2^2) = [0, 0.2]$, $S^{\rm in}(\mathbf{x}_2^2) = [-0.4, 0.2]$ for DNN $\mathcal{N}_e$, and by interval analysis, we have $S(\hat{\mathbf{x}}_1^1) = S^{\rm in}(\hat{\mathbf{x}}_1^1) = [6, 12]$, $S(\hat{\mathbf{x}}_2^1) = S^{\rm in}(\hat{\mathbf{x}}_2^1) = [3, 9]$, $S(\hat{\mathbf{x}}_1^2) = S^{\rm in}(\hat{\mathbf{x}}_1^2) = [1, 4]$, $S(\hat{\mathbf{x}}_2^2) = [0, 1]$, $S^{\rm in}(\hat{\mathbf{x}}_2^2) = [-2, 1]$ for QNN $\widehat{\mathcal{N}}_e$. According to Definition 3 and Algorithm 1, we then have $\delta_{1,1} = [-0.05, -0.025]$, $\delta_{1,2} = [-0.0375, -0.0125]$, $\Delta\mathbf{b}_1^2 = \Delta\mathbf{b}_2^2 = 0$, and rounding error $\xi = 0.125$.

Then, we get the input difference intervals $\delta_{2,1}^{in}$ and $\delta_{2,2}^{in}$ as follows based on Algorithm 2:

- $\mathrm{LB}(\delta_{2,1}^{in}) = 1.25 \times \mathrm{LB}(\delta_{1,1}) + (1.25 - 1.2) \times \mathrm{LB}(S(\mathbf{x}_1^1)) + (-0.25) \times \mathrm{UB}(\delta_{1,2}) + (-0.25 + 0.2) \times \mathrm{UB}(S(\mathbf{x}_2^1)) - \xi = -0.194375$;
- $\mathrm{UB}(\delta_{2,1}^{in}) = 1.25 \times \mathrm{UB}(\delta_{1,1}) + (1.25 - 1.2) \times \mathrm{UB}(S(\mathbf{x}_1^1)) + (-0.25) \times \mathrm{LB}(\delta_{1,2}) + (-0.25 + 0.2) \times \mathrm{LB}(S(\mathbf{x}_2^1)) + \xi = 0.133125$;
- $\mathrm{LB}(\delta_{2,2}^{in}) = -0.75 \times \mathrm{UB}(\delta_{1,1}) + (-0.75 + 0.7) \times \mathrm{UB}(S(\mathbf{x}_1^1)) + 0.75 \times \mathrm{LB}(\delta_{1,2}) + (0.75 - 0.8) \times \mathrm{UB}(S(\mathbf{x}_2^1)) - \xi = -0.204375$;
- $\mathrm{UB}(\delta_{2,2}^{in}) = -0.75 \times \mathrm{LB}(\delta_{1,1}) + (-0.75 + 0.7) \times \mathrm{LB}(S(\mathbf{x}_1^1)) + 0.75 \times \mathrm{UB}(\delta_{1,2}) + (0.75 - 0.8) \times \mathrm{LB}(S(\mathbf{x}_2^1)) + \xi = 0.123125$;

According to Algorithm 3, we have

- $\delta_{2,1}^{in} = \delta_{2,1} = [-0.194375, 0.133125]$;
- $\delta_{2,2} = [-0.2, 0.123125]$:
  - $\mathrm{LB}(\delta_{2,2})=\max\big(\mathrm{LB}(\delta_{2,2}^{in}), -\mathrm{UB}(S^{in}(\mathbf{x}_2^2))\big) = \max(-0.204375, -0.2) = -0.2$;
  - $\mathrm{UB}(\delta_{2,2})=\min\big(\mathrm{UB}(\delta_{2,2}^{in}), \mathrm{UB}(S^{in}(\tilde{\mathbf{x}}_2^2))\big) = \min(0.123125, 0.25) = 0.123125$.

Therefore, we get the input/output difference interval $\delta_{3,1} = \delta_{3,1}^{in}$ as follows:

- $\mathrm{LB}(\delta_{3,1}^{in}) = 0.25 \times \mathrm{LB}(\delta_{2,1}) + (0.25 - 0.3) \times \mathrm{UB}(S(\mathbf{x}_1^2)) + 0.75 \times \mathrm{LB}(\delta_{2,2}) + (0.75 - 0.7) \times \mathrm{LB}(S(\mathbf{x}_2^2)) = -0.24459375$;
- $\mathrm{UB}(\delta_{3,1}^{in}) = 0.25 \times \mathrm{UB}(\delta_{2,1}) + (0.25 - 0.3) \times \mathrm{LB}(S(\mathbf{x}_1^2)) + 0.75 \times \mathrm{UB}(\delta_{2,2}) + (0.75 - 0.7) \times \mathrm{UB}(S(\mathbf{x}_2^2)) = 0.117625$;

Finally, the quantization error interval is $[-0.24459375, 0.117625]$.

2.2 Example for Section 4

Example 2. Consider the DNN $\mathcal{N}_e$ and QNN $\widehat{\mathcal{N}}_e$ given in Fig. 1(a) and Fig. 1(b) again, where $2^{-F_{in}} = 1/16$, $\mathcal{C}_h^{\mathrm{ub}} = 1/15$. We set the input region for QNN $\widehat{\mathcal{N}}_e$ as $R((9,6),3) = \{(x,y) \in \mathbb{Z}^2 \mid 6 \le x \le 12, 3 \le y \le 9\}$.

Different from example 1, we use symbolic-based method given in Section 4 to compute the intervals $\delta_{2,1}^{in}$ and $\delta_{2,2}^{in}$.

Firstly, we get the abstract element $\mathcal{A}_{j,s}^2 = \langle \mathbf{a}_{j,s}^{2,\le}, \mathbf{a}_{j,s}^{2,\ge}, l_{j,s}^2, u_{j,s}^2 \rangle$ for $j \in \{1,2\}$ and $s \in \{0,1\}$ for DNN $\mathcal{N}_e$ as follows:

- $\mathbf{x}_1^2 \to \mathbf{x}_{1,0}^2, \mathbf{x}_{1,1}^2$:
  - $\mathcal{A}_{1,0}^2 = \langle 1.2\mathbf{x}_1^1 - 0.2\mathbf{x}_2^1,\ 1.2\mathbf{x}_1^1 - 0.2\mathbf{x}_2^1,\ 0.36,\ 0.92 \rangle$;
  - $\mathcal{A}_{1,1}^2 = \langle \mathbf{x}_{1,0}^2,\ \mathbf{x}_{1,0}^2,\ 0.36,\ 0.92 \rangle$.
- $\mathbf{x}_2^2 \to \mathbf{x}_{2,0}^2, \mathbf{x}_{2,1}^2$:
  - $\mathcal{A}_{2,0}^2 = \langle -0.7\mathbf{x}_1^1 + 0.8\mathbf{x}_2^1,\ -0.7\mathbf{x}_1^1 + 0.8\mathbf{x}_2^1,\ -0.4,\ 0.2 \rangle$;
  - $\mathcal{A}_{2,1}^2 = \langle 0,\ \frac{1}{3}\mathbf{x}_{2,0}^2 + \frac{0.4}{3},\ 0,\ 0.2 \rangle$.

After substituting every variable in $\mathbf{a}_{1,1}^{2,\le}$, $\mathbf{a}_{1,1}^{2,\ge}$, $\mathbf{a}_{2,1}^{2,\le}$, and $\mathbf{a}_{2,1}^{2,\ge}$ until no further substitution is possible, we have the following forms of linear combination of the input variables:

- $\mathcal{A}_{1,0}^{2,*} = \langle 1.2\mathbf{x}_1^1 - 0.2\mathbf{x}_2^1,\ 1.2\mathbf{x}_1^1 - 0.2\mathbf{x}_2^1,\ 0.36,\ 0.92 \rangle$;
- $\mathcal{A}_{1,1}^{2,*} = \langle 1.2\mathbf{x}_1^1 - 0.2\mathbf{x}_2^1,\ 1.2\mathbf{x}_1^1 - 0.2\mathbf{x}_2^1,\ 0.36,\ 0.92 \rangle$;
- $\mathcal{A}_{2,0}^{2,*} = \langle -0.7\mathbf{x}_1^1 + 0.8\mathbf{x}_2^1,\ -0.7\mathbf{x}_1^1 + 0.8\mathbf{x}_2^1,\ -0.4,\ 0.2 \rangle$;
- $\mathcal{A}_{2,1}^{2,*} = \langle 0,\ -\frac{0.7}{3}\mathbf{x}_1^1 + \frac{0.8}{3}\mathbf{x}_2^1 + \frac{0.4}{3},\ 0,\ 0.2 \rangle$.

Then, we get the abstract element $\widehat{\mathcal{A}}_{j,p}^2 = \langle \hat{\mathbf{a}}_{j,p}^{2,\le}, \hat{\mathbf{a}}_{j,p}^{2,\ge}, \hat{l}_{j,p}^2, \hat{u}_{j,p}^2 \rangle$ for $j \in \{1,2\}$ and $p \in \{0,1,2\}$ for QNN $\widehat{\mathcal{N}}_e$ as follows:

- $\hat{\mathbf{x}}_1^2 \to \hat{\mathbf{x}}_{1,0}^2, \hat{\mathbf{x}}_{1,1}^2, \hat{\mathbf{x}}_{1,2}^2$:
  - $\widehat{\mathcal{A}}_{1,0}^2 = \langle \frac{1}{16}(5\hat{\mathbf{x}}_1^1 - \hat{\mathbf{x}}_2^1) - 0.5,\ \frac{1}{16}(5\hat{\mathbf{x}}_1^1 - \hat{\mathbf{x}}_2^1) + 0.5,\ 0.8125,\ 3.0625 \rangle$;
  - $\widehat{\mathcal{A}}_{1,1}^2 = \langle \hat{\mathbf{x}}_{1,0}^2,\ \hat{\mathbf{x}}_{1,0}^2,\ 0.8125,\ 3.0625 \rangle$;
  - $\widehat{\mathcal{A}}_{1,2}^2 = \langle \hat{\mathbf{x}}_{1,1}^2,\ \hat{\mathbf{x}}_{1,1}^2,\ 0.8125,\ 3.0625 \rangle$.
- $\hat{\mathbf{x}}_2^2 \to \hat{\mathbf{x}}_{2,0}^2, \hat{\mathbf{x}}_{2,1}^2, \hat{\mathbf{x}}_{2,2}^2$:
  - $\widehat{\mathcal{A}}_{2,0}^2 = \langle \frac{1}{16}(-3\hat{\mathbf{x}}_1^1 + 3\hat{\mathbf{x}}_2^1) - 0.5,\ \frac{1}{16}(-3\hat{\mathbf{x}}_1^1 + 3\hat{\mathbf{x}}_2^1) + 0.5,\ -2.1875,\ 1.0625 \rangle$;
  - $\widehat{\mathcal{A}}_{2,1}^2 = \langle 0, \frac{17(\hat{\mathbf{x}}_{2,0}^2 + 2.1875)}{52}, 0, 1.0625 \rangle$;
  - $\widehat{\mathcal{A}}_{2,2}^2 = \langle \hat{\mathbf{x}}_{2,1}^2,\ \hat{\mathbf{x}}_{2,1}^2,\ 0,\ 1.0625 \rangle$.

After substituting every variable in $\hat{\mathbf{a}}_{1,1}^{2,\le}$, $\hat{\mathbf{a}}_{1,1}^{2,\ge}$, $\hat{\mathbf{a}}_{1,2}^{2,\le}$, $\hat{\mathbf{a}}_{1,2}^{2,\ge}$, $\hat{\mathbf{a}}_{2,1}^{2,\le}$, $\hat{\mathbf{a}}_{2,1}^{2,\ge}$, $\hat{\mathbf{a}}_{2,2}^{2,\le}$, and $\hat{\mathbf{a}}_{2,2}^{2,\ge}$ until no further substitution is possible, we have the following forms of linear combination of the input variables:

- $\widehat{\mathcal{A}}_{1,0}^{2,*} = \langle \frac{1}{16}(5\hat{\mathbf{x}}_1^1 - \hat{\mathbf{x}}_2^1) - 0.5,\ \frac{1}{16}(5\hat{\mathbf{x}}_1^1 - \hat{\mathbf{x}}_2^1) + 0.5,\ 0.8125,\ 3.0625 \rangle$;
- $\widehat{\mathcal{A}}_{1,1}^{2,*} = \langle \frac{1}{16}(5\hat{\mathbf{x}}_1^1 - \hat{\mathbf{x}}_2^1) - 0.5,\ \frac{1}{16}(5\hat{\mathbf{x}}_1^1 - \hat{\mathbf{x}}_2^1) + 0.5,\ 0.8125,\ 3.0625 \rangle$;

- $\widehat{\mathcal{A}}^{2,*}_{1,2} = \langle \frac{1}{16}(5\hat{\mathbf{x}}^1_1 - \hat{\mathbf{x}}^1_2) - 0.5,\ \frac{1}{16}(5\hat{\mathbf{x}}^1_1 - \hat{\mathbf{x}}^1_2) + 0.5,\ 0.8125,\ 3.0625\rangle$;
- $\widehat{\mathcal{A}}^{2,*}_{2,0} = \langle \frac{1}{16}(-3\hat{\mathbf{x}}^1_1 + 3\hat{\mathbf{x}}^1_2) - 0.5,\ \frac{1}{16}(-3\hat{\mathbf{x}}^1_1 + 3\hat{\mathbf{x}}^1_2) + 0.5,\ -2.1875,\ 1.0625\rangle$;
- $\widehat{\mathcal{A}}^{2,*}_{2,1} = \langle 0,\ \frac{\frac{17}{16}(-3\hat{\mathbf{x}}^1_1+3\hat{\mathbf{x}}^1_2)+45.6875}{52},\ 0,\ 1.0625\rangle$;
- $\widehat{\mathcal{A}}^{2,*}_{2,2} = \langle 0,\ \frac{\frac{17}{16}(-3\hat{\mathbf{x}}^1_1+3\hat{\mathbf{x}}^1_2)+45.6875}{52},\ 0,\ 1.0625\rangle$;

Therefore, we have the lower bounds $\Delta l^{2,*}_{1,0}$, $\Delta l^{2,*}_{2,0}$, and upper bounds $\Delta u^{2,*}_{1,0}$, $\Delta u^{2,*}_{2,0}$ of the difference interval $\delta^{in}_{2,1}$, $\delta^{in}_{2,2}$ for the hidden neurons based on the input region as well as $\hat{\mathbf{x}}^1_j = 15\mathbf{x}^1_j$ for $j \in \{1,2\}$ as follows:

- $\Delta l^{2,*}_{1,0} = 2^{-2}(\hat{\mathbf{a}}^{2,\leq,*}_{1,0}) - \mathbf{a}^{2,\geq,*}_{1,0} = 2^{-2}\big(\frac{1}{16}(5\hat{\mathbf{x}}^1_1 - \hat{\mathbf{x}}^1_2) - 0.5\big) - (1.2\mathbf{x}^1_1 - 0.2\mathbf{x}^1_2)$, and $\mathrm{LB}(\Delta l^{2,*}_{1,0}) = -0.168125$;
- $\Delta u^{2,*}_{1,0} = 2^{-2}(\hat{\mathbf{a}}^{2,\geq,*}_{1,0}) - \mathbf{a}^{2,\leq,*}_{1,0} = 2^{-2}\big(\frac{1}{16}(5\hat{\mathbf{x}}^1_1 - \hat{\mathbf{x}}^1_2) + 0.5\big) - (1.2\mathbf{x}^1_1 - 0.2\mathbf{x}^1_2)$, and $\mathrm{UB}(\Delta u^{2,*}_{1,0}) = 0.081875$.
- $\Delta l^{2,*}_{2,0} = 2^{-2}(\hat{\mathbf{a}}^{2,\leq,*}_{2,0}) - \mathbf{a}^{2,\geq,*}_{2,0} = 2^{-2}\big(\frac{1}{16}(-3\hat{\mathbf{x}}^1_1 + 3\hat{\mathbf{x}}^1_2) - 0.5\big) - (-0.7\mathbf{x}^1_1 + 0.8\mathbf{x}^1_2)$, and $\mathrm{LB}(\Delta l^{2,*}_{2,0}) = -0.185625$;
- $\Delta u^{2,*}_{2,0} = 2^{-2}(\hat{\mathbf{a}}^{2,\geq,*}_{2,0}) - \mathbf{a}^{2,\leq,*}_{2,0} = 2^{-2}\big(\frac{1}{16}(-3\hat{\mathbf{x}}^1_1 + 3\hat{\mathbf{x}}^1_2) + 0.5\big) - (-0.7\mathbf{x}^1_1 + 0.8\mathbf{x}^1_2)$, and $\mathrm{UB}(\Delta u^{2,*}_{2,0}) = 0.104375$.

Note that, based on above, we can compute $S^{in}(\hat{\mathbf{x}}^2_1) = [0.8125, 3.0625]$ and $S^{in}(\hat{\mathbf{x}}^2_1) = [-2.1875, 1.0625]$ via symbolic interval analysis on QNN $\widehat{\mathcal{N}}_e$. Then, according to Algorithm 3, we have:

- $\delta^{in}_{2,1} = \delta_{2,1} = [-0.168125, 0.081875]$;
- $\delta_{2,2} = [-0.185625, 0.104375]$:
  - $\mathrm{LB}(\delta_{2,2}) = \max\big(\mathrm{LB}(\delta^{in}_{2,2}), -\mathrm{UB}(S^{in}(\mathbf{x}^2_2))\big) = \max(-0.185625, -0.2) = -0.185625$;
  - $\mathrm{UB}(\delta_{2,2}) = \min\big(\mathrm{UB}(\delta^{in}_{2,2}), \mathrm{UB}(S^{in}(\tilde{\mathbf{x}}^2_2))\big) = \min(0.104375, 1.0625/4) = 0.104375$.

We remark that for the output layers in DNN $\mathcal{N}_e$ and QNN $\widehat{\mathcal{N}}_e$, we also have $\mathbf{x}^3_1 = \mathbf{x}^3_{1,0} = \mathbf{x}^2_{1,1} + 3\mathbf{x}^2_{2,1}$ and $\hat{\mathbf{x}}^3_1 = \hat{\mathbf{x}}^3_{1,0} = \hat{\mathbf{x}}^2_{1,2} + 3\hat{\mathbf{x}}^2_{2,2}$. Hence, for the output layer, we have:

- $\mathbf{a}^{3,\leq,*}_{1,0} = 0.3 \times \mathbf{a}^{2,\leq,*}_{1,1} + 0.7 \times \mathbf{a}^{2,\leq,*}_{2,1}$, $\mathbf{a}^{3,\geq,*}_{1,0} = 0.3 \times \mathbf{a}^{2,\geq,*}_{1,1} + 0.7 \times \mathbf{a}^{2,\geq,*}_{2,1}$;
- $\hat{\mathbf{a}}^{3,\leq,*}_{1,0} = 2^{-2}(\hat{\mathbf{a}}^{2,\leq,*}_{1,2} + 3 \times \hat{\mathbf{a}}^{2,\leq,*}_{2,2})$, $\hat{\mathbf{a}}^{3,\geq,*}_{1,0} = 2^{-2}(\hat{\mathbf{a}}^{2,\geq,*}_{1,2} + 3 \times \hat{\mathbf{a}}^{2,\geq,*}_{2,2})$.

Finally, we get the lower bound $\Delta l^{3,*}_{1,0}$ and upper bound $\Delta u^{3,*}_{1,0}$ for the output neurons as follows:

- $\Delta l^{3,*}_{1,0} = 2^{-2}(\hat{\mathbf{a}}^{3,\leq,*}_{1,0}) - \mathbf{a}^{3,\geq,*}_{1,0} = 2^{-4}(\hat{\mathbf{a}}^{2,\leq,*}_{1,2} + 3 \times \hat{\mathbf{a}}^{2,\leq,*}_{2,2}) - (0.3 \times \mathbf{a}^{2,\geq,*}_{1,1} + 0.7 \times \mathbf{a}^{2,\geq,*}_{2,1})$, and $\mathrm{LB}(\Delta l^{3,*}_{1,0}) = -0.197219$;
- $\Delta u^{3,*}_{1,0} = 2^{-2}(\hat{\mathbf{a}}^{3,\geq,*}_{1,0}) - \mathbf{a}^{3,\leq,*}_{1,0} = 2^{-4}(\hat{\mathbf{a}}^{2,\geq,*}_{1,2} + 3 \times \hat{\mathbf{a}}^{2,\geq,*}_{2,2}) - (0.3 \times \mathbf{a}^{2,\leq,*}_{1,1} + 0.7 \times \mathbf{a}^{2,\leq,*}_{2,1})$, and $\mathrm{UB}(\Delta u^{3,*}_{1,0}) = 0.2045$.

2.3 Analysis of Results

Note that, although in the Examples 1 and 2, the quantization error interval computed by QEBVerif (sym) is looser than that by QEBVerif (Con), QEBVerif performs better on computing the difference intervals for the hidden neurons, and such a comparison result is quite similar to the cases of P1-8 and P1-10 in Table 2. However, we remark that QEBVerif (sym) works better in most cases, especially when there are more than 1 hidden layer, as shown in Table 2.