Source task and conventions

Specify all hidden and output weights and offsets for the source sign convention.

Conventional solution and proof

For $x_1, x_2 \in \{0, 1\}$, take

$$z_1 = \mathrm{ReLU}(x_1 - x_2), \qquad z_2 = \mathrm{ReLU}(x_2 - x_1), \qquad y = \mathrm{Sign}(z_1 + z_2 - 1/2).$$

The hidden weight rows are $(1, -1)$ and $(-1, 1)$, both hidden biases are zero, the output weights are $(1, 1)$, and the output bias is $-1/2$. For inputs $(0, 0), (1, 0), (0, 1), (1, 1)$ the hidden pairs are respectively $(0, 0), (1, 0), (0, 1), (0, 0)$; the output preactivations are $-1/2, 1/2, 1/2, -1/2$. Applying the stated sign convention gives exactly $-1, 1, 1, -1$, with no boundary ambiguity. Two hidden units are also minimal for a network of the specified form without skip connections. With one ReLU unit, the sign of $a\,\mathrm{ReLU}(w^\top x + b) + c$ defines a halfspace, its complement, or a constant set along the scalar $w^\top x + b$. XOR is not linearly separable: if an affine function is positive at $(1, 0), (0, 1)$ and nonpositive at $(0, 0), (1, 1)$, adding the first two inequalities contradicts the latter two. Therefore one hidden unit cannot realize this truth table.

Structural or analytic derivation

The hidden sum equals $|x_1 - x_2|$, so the network tests whether the two Boolean inputs differ. On the Boolean domain the same function is the polynomial $2x_1 + 2x_2 - 4x_1x_2 - 1$. That polynomial is a verification identity for the four inputs; it is not a substitution for the required ReLU architecture off the Boolean domain.

Comparative cost and preferred method