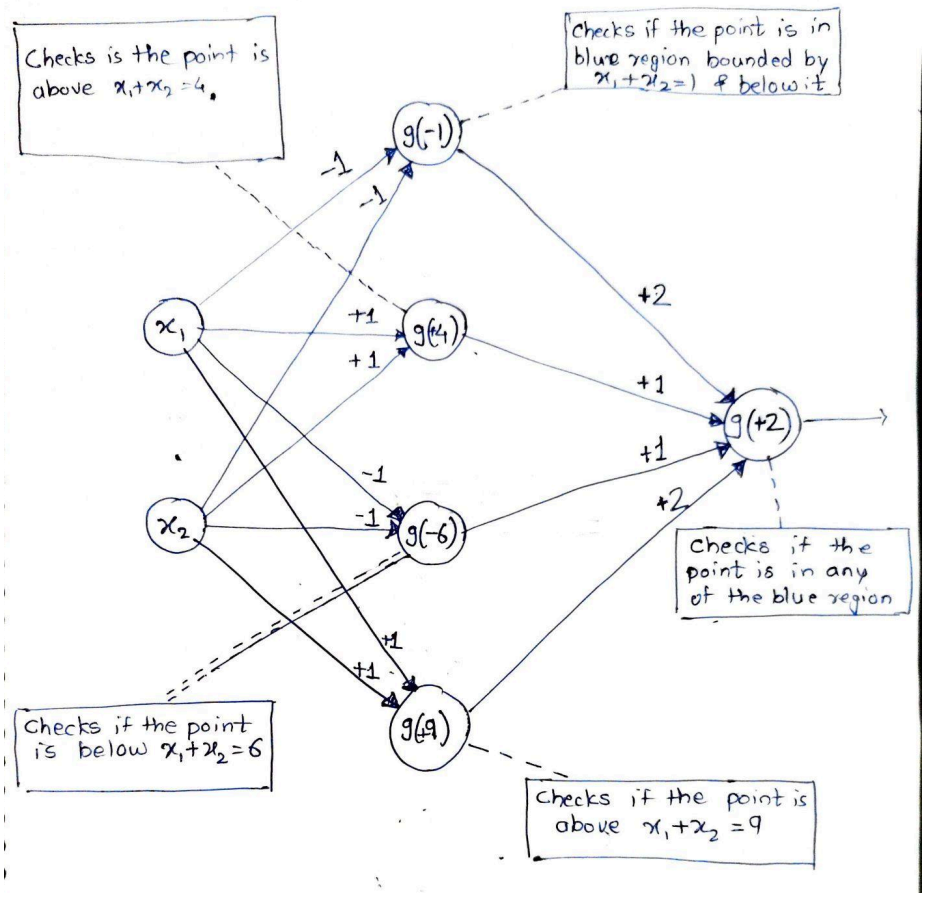
BANDS OF BLUE:
Let g be the step activation function which we will use as the default activation function for all the neurons. Formally, it's defined as

$$g(x;T) = \begin{cases} 1 & \text{If } x \geq T \\ 0 & \text{Otherwise} \end{cases}$$

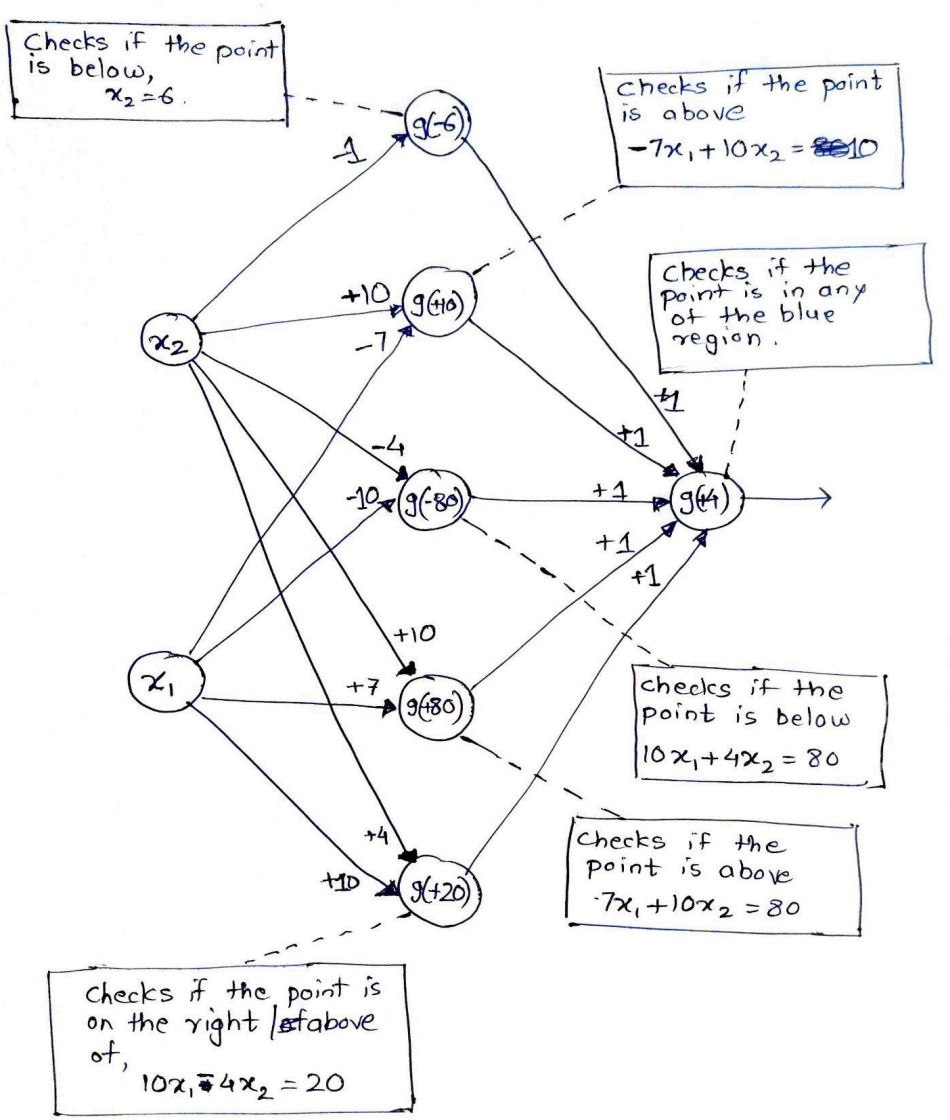
Then the optimal neural network for the dataset specified above looks as follows:



Hidden Layer 1: Geometrically, these neurons are determining whether the point is lying above or below the lines. When the weights are negative we are testing if the point is below the line and vice versa.

Output layer: This checks if the point is in any of the blue regions and outputs 1 if true.

Catch the star



Hidden Layer 1: Geometrically, these neurons are determining whether the point is lying above or below the lines. When the weights are negative we are testing if the point is below the line and vice versa.

Output layer: This checks if the point is in any of the blue regions and outputs 1 if true. The sum of outputs from all the layers turns out to be at least 4 (when the point is in blue triangular region) and 5 (if the point is in central pentagonal region).