Q1. Neural Networks: Representation

G₁:

G₂:

G₃:

G₄:

G₅:

H₁:

H₂:

H₃:

H₄:

H₅:
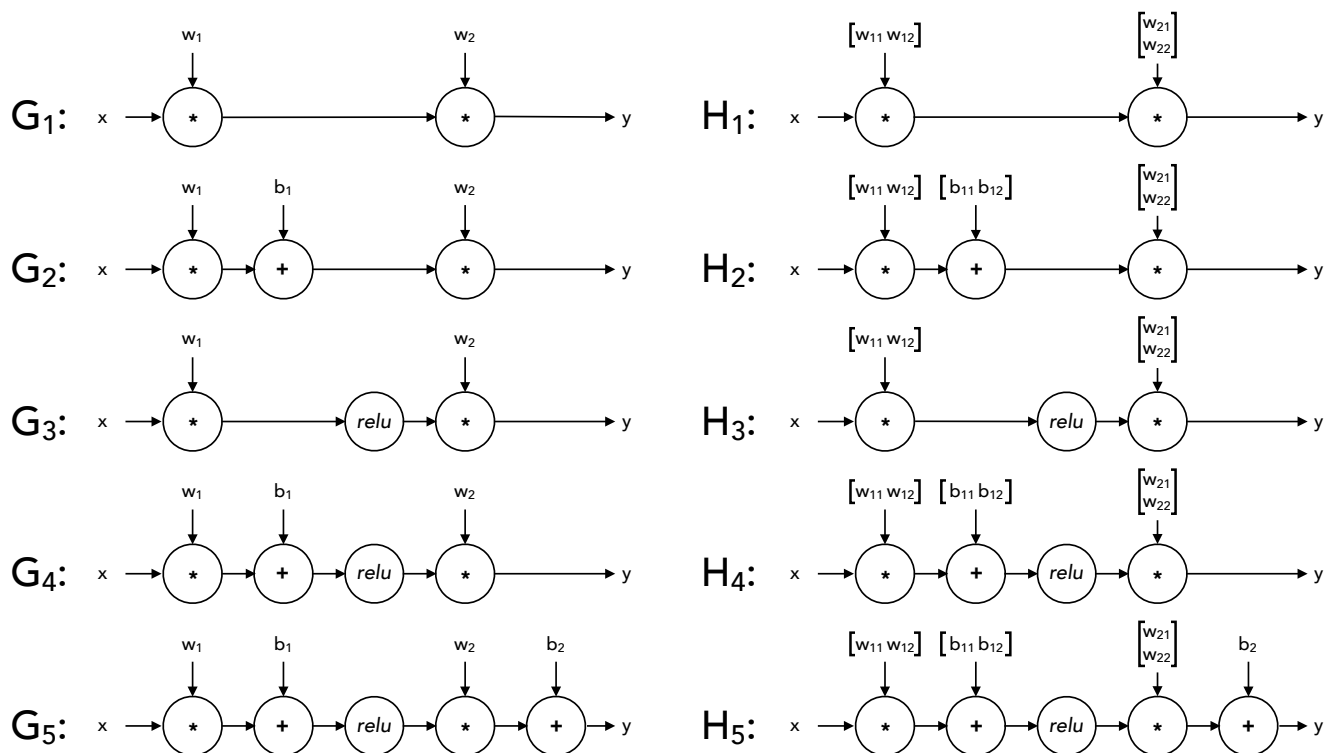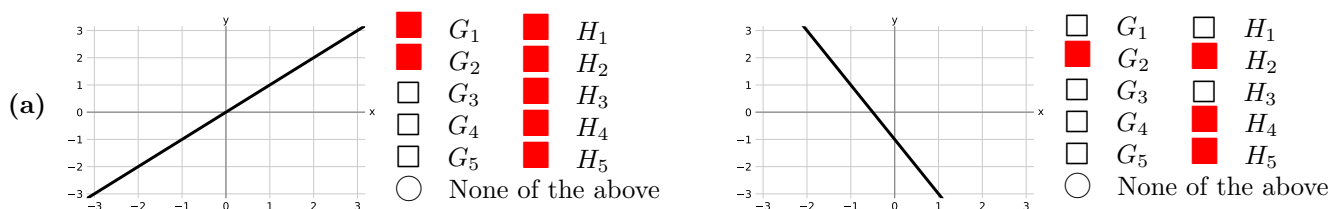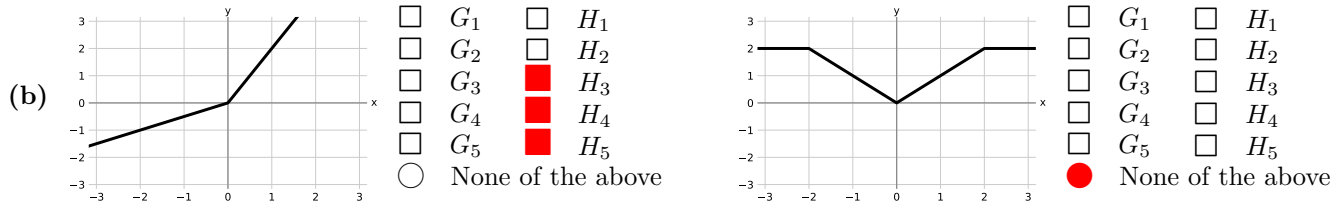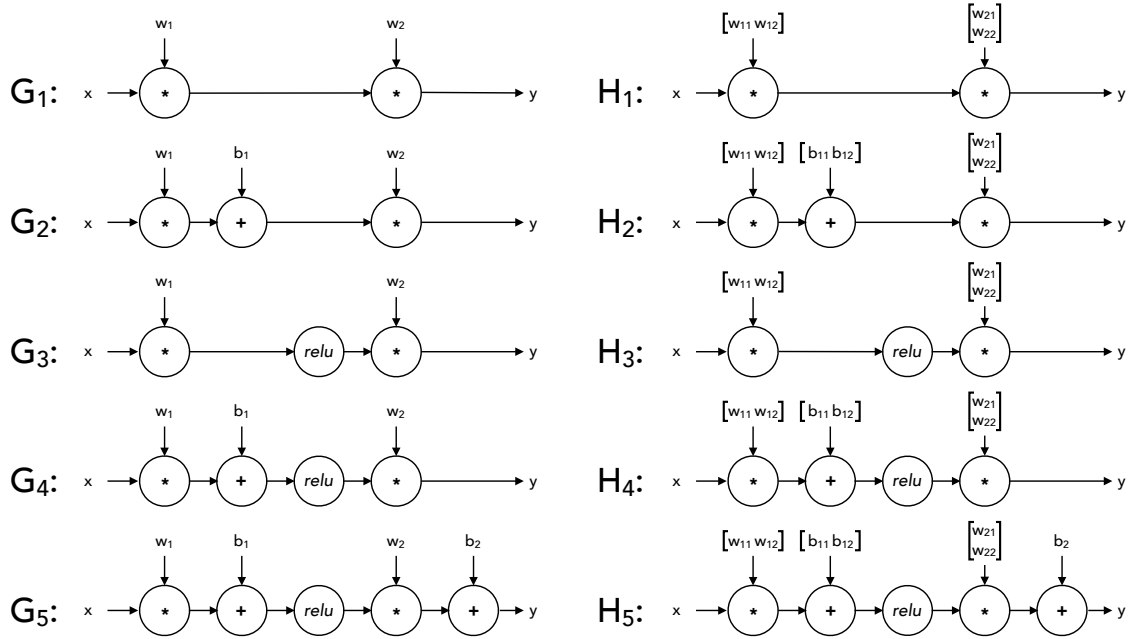
For each of the piecewise-linear functions below, mark all networks from the list above that can represent the function **exactly** on the range $x \in (-\infty, \infty)$. In the networks above, *relu* denotes the element-wise ReLU nonlinearity: $relu(z) = max(0, z)$. The networks G_i use 1-dimensional layers, while the networks H_i have some 2-dimensional intermediate layers.

(a)

■	G_1	■	H_1
■	G_2	■	H_2
☐	G_3	■	H_3
☐	G_4	■	H_4
☐	G_5	■	H_5
○	None of the above		

☐	G_1	☐	H_1
■	G_2	■	H_2
☐	G_3	☐	H_3
☐	G_4	■	H_4
☐	G_5	■	H_5
○	None of the above		

The networks G_3, G_4, G_5 include a ReLU nonlinearity on a scalar quantity, so it is impossible for their output to represent a non-horizontal straight line. On the other hand, H_3, H_4, H_5 have a 2-dimensional hidden layer, which allows two ReLU elements facing in opposite directions to be added together to form a straight line. The second subpart requires a bias term because the line does not pass through the origin.
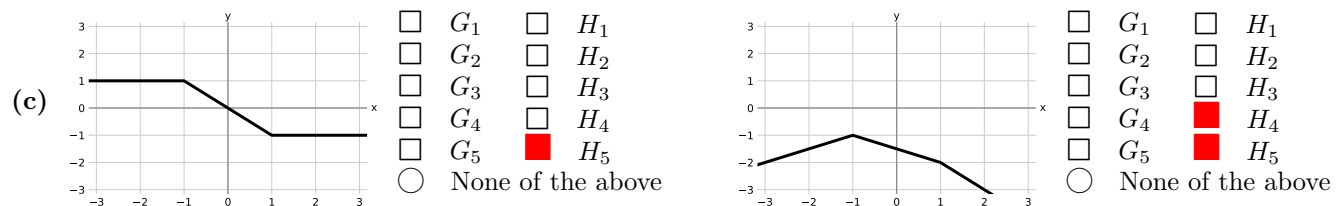
The list of networks is reproduced below for your convenience:

(b)

☐	G_1	☐	H_1
☐	G_2	☐	H_2
☐	G_3	■	H_3
☐	G_4	■	H_4
☐	G_5	■	H_5
○	None of the above		

☐	G_1	☐	H_1
☐	G_2	☐	H_2
☐	G_3	☐	H_3
☐	G_4	☐	H_4
☐	G_5	☐	H_5
●	None of the above		

These functions include multiple non-horizontal linear regions, so they cannot be represented by any of the networks G_i which apply ReLU no more than once to a scalar quantity.

The first subpart can be represented by any of the networks with 2-dimensional ReLU nodes. The point of nonlinearity occurs at the origin, so nonzero bias terms are not required.

The second subpart has 3 points where the slope changes, but the networks H_i only have a single 2-dimensional ReLU node. Each application of ReLU to one element can only introduce a change of slope for a single value of x.

(c)

☐	G_1	☐	H_1
☐	G_2	☐	H_2
☐	G_3	☐	H_3
☐	G_4	☐	H_4
☐	G_5	■	H_5
○	None of the above		

☐	G_1	☐	H_1
☐	G_2	☐	H_2
☐	G_3	☐	H_3
☐	G_4	■	H_4
☐	G_5	■	H_5
○	None of the above		

Both functions have two points where the slope changes, so none of the networks $G_i; H_1, H_2$ can represent them.

An output bias term is required for the first subpart because one of the flat regions must be generated by the flat part of a ReLU function, but neither one of them is at $y = 0$.

The second subpart doesn't require a bias term at the output: it can be represented as $-relu(\frac{-x+1}{2}) - relu(x+1)$. Note how if the segment at $x > 2$ were to be extended to cross the x axis, it would cross exactly at $x = -1$, the location of the other slope change. A similar statement is true for the segment at $x < -1$.

Q2. Decision Trees

You are given points from 2 classes, shown as +'s and ·'s. For each of the following sets of points,

1. Draw the decision tree of depth at most 2 that can separate the given data completely, by filling in binary predicates (which only involve thresholding of a *single* variable) in the boxes for the decision trees below. If the data is already separated when you hit a box, simply write the class, and leave the sub-tree hanging from that box empty.

2. Draw the corresponding decision boundaries on the scatter plot, and write the class labels for each of the resulting bins somewhere inside the resulting bins.

If the data can not be separated completely by a depth 2 decision tree, simply cross out the tree template. We solve the first part as an example.