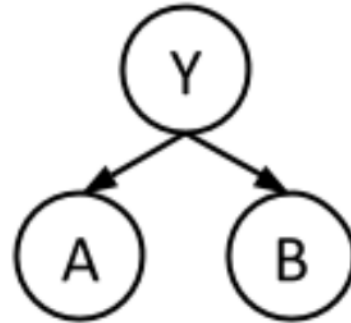


# 1 Naive Bayes

In this question, we will train a Naive Bayes classifier to predict class labels  $Y$  as a function of input features  $A$  and  $B$ .  $Y$ ,  $A$ , and  $B$  are all binary variables, with domains 0 and 1. We are given 10 training points from which we will estimate our distribution.

$A$	1	1	1	1	0	1	0	1	1	1
$B$	1	0	0	1	1	1	1	0	1	1
$Y$	1	1	0	0	0	1	1	0	0	0



1. What are the maximum likelihood estimates for the tables  $P(Y)$ ,  $P(A|Y)$ , and  $P(B|Y)$ ?

$Y$	$P(Y)$
0	
1	

$A$	$Y$	$P(A Y)$
0	0	
1	0	
0	1	
1	1	

$B$	$Y$	$P(B Y)$
0	0	
1	0	
0	1	
1	1	

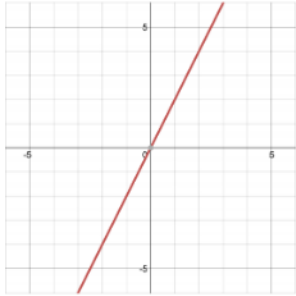
2. Consider a new data point ( $A = 1, B = 1$ ). What label would this classifier assign to this sample?

3. Let's use Laplace Smoothing to smooth out our distribution. Compute the new distribution for  $P(A|Y)$  given Laplace Smoothing with  $k = 2$ .

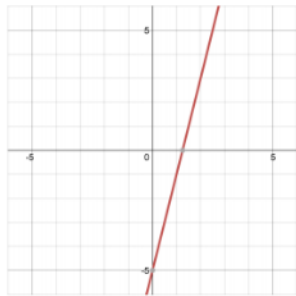
$A$	$Y$	$P(A Y)$
0	0	
1	0	
0	1	
1	1	

## 2 Neural Network Representations

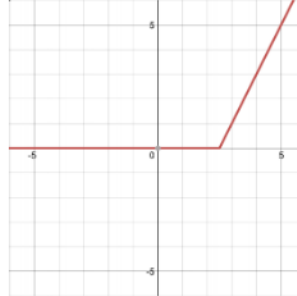
You are given a number of functions (a-h) of a single variable,  $x$ , which are graphed below. The computation graphs on the following pages will start off simple and get more complex, building up to neural networks. For each computation graph, indicate which of the functions below they are able to represent.



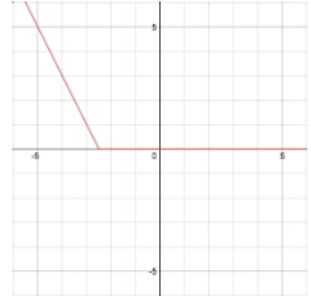
(a)  $2x$



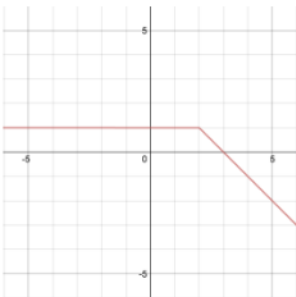
(b)  $4x - 5$



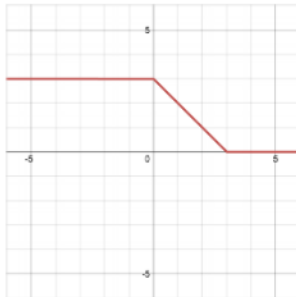
(c)  $\begin{cases} 2x - 5 & x \geq 2.5 \\ 0 & x < 2.5 \end{cases}$



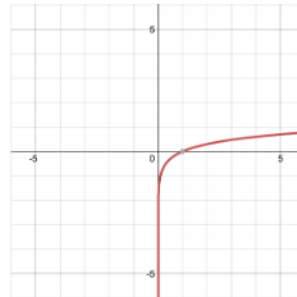
(d)  $\begin{cases} -2x - 5 & x \leq -2.5 \\ 0 & x > -2.5 \end{cases}$



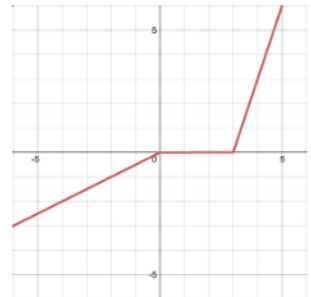
(e)  $\begin{cases} -x + 3 & x \geq 2 \\ 1 & x < 2 \end{cases}$



(f)  $\begin{cases} 3 & x \leq 0 \\ 3 - x & 0 < x \leq 3 \\ 0 & x > 3 \end{cases}$

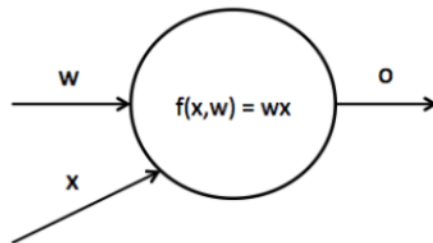


(g)  $\log(x)$

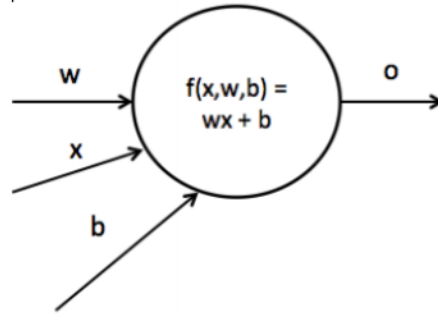


(h)  $\begin{cases} 0.5x & x \leq 0 \\ 0 & 0 < x \leq 3 \\ 3x - 9 & x > 3 \end{cases}$

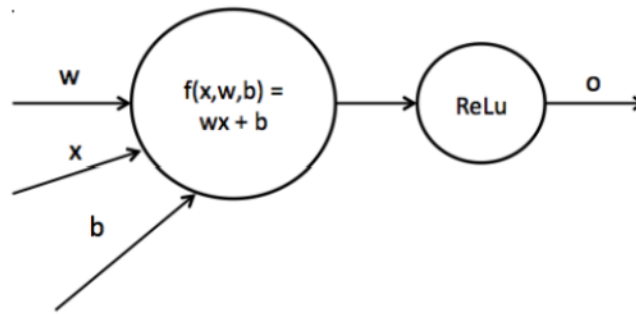
1. Consider the following computation graph, computing a linear transformation with scalar input  $x$ , weight  $w$ , and output  $o$ , such that  $o = wx$ . Which of the functions can be represented by this graph? For the options which can, write out the appropriate value of  $w$ .



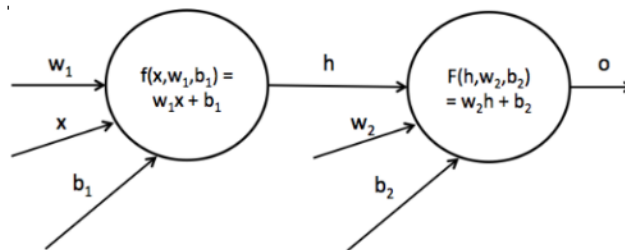
2. Now we introduce a bias term  $b$  into the graph, such that  $o = wx + b$  (this is known as an *affine* function). Which of the functions can be represented by this network? For the options which can, write out an appropriate value of  $w, b$ .



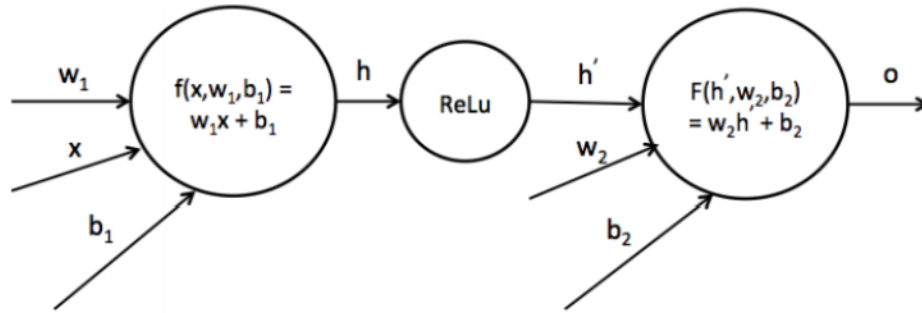
3. We can introduce a non-linearity into the network as indicated below. We use the ReLU non-linearity, which has the form  $ReLU(x) = \max(0, x)$ . Now which of the functions can be represented by this neural network with weight  $w$  and bias  $b$ ? For the options which can, write out an appropriate value of  $w, b$ .



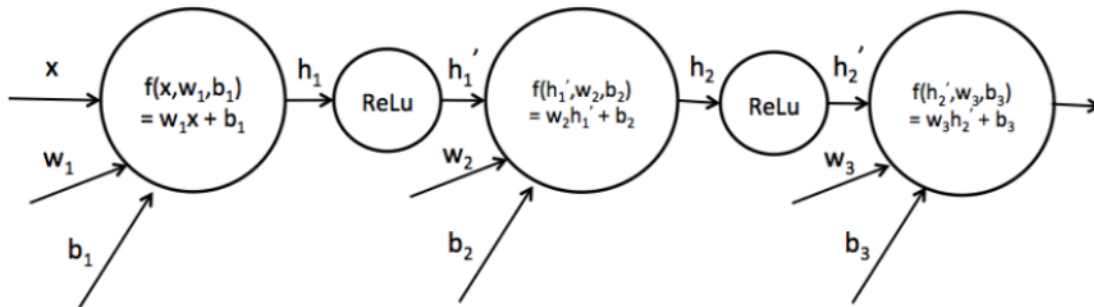
4. Now we consider neural networks with multiple affine transformations, as indicated below. We now have two sets of weights and biases  $w_1, b_1$  and  $w_2, b_2$ . We denote the result of the first transformation  $h$  such that  $h = w_1x + b_1$ , and  $o = w_2h + b_2$ . Which of the functions can be represented by this network? For the options which can, write out appropriate values of  $w_1, w_2, b_1, b_2$ .



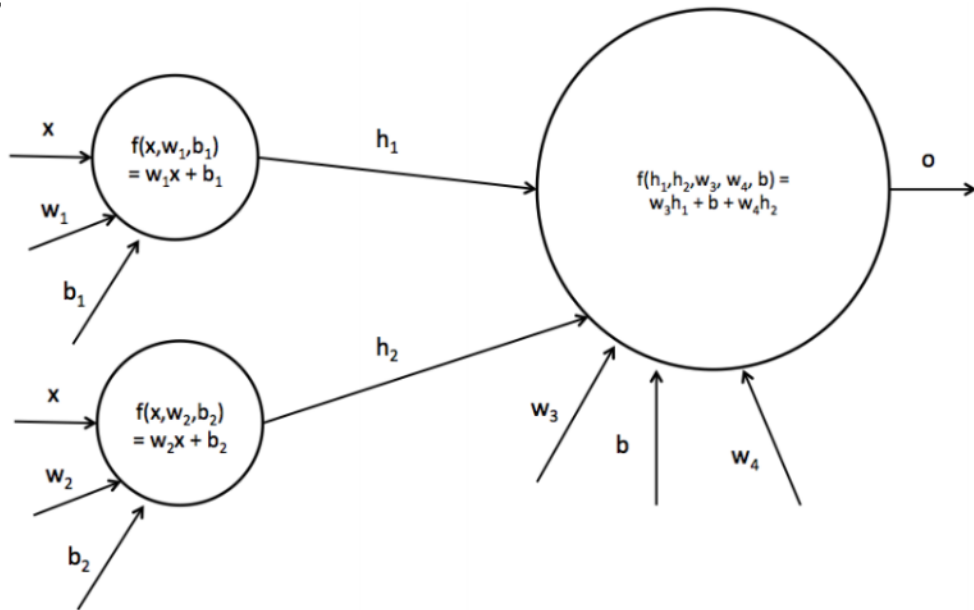
5. Next we add a ReLU non-linearity to the network after the first affine transformation, creating a hidden layer. Which of the functions can be represented by this network? For the options which can, write out appropriate values of  $w_1, w_2, b_1, b_2$ .



6. Now we add another hidden layer to the network, as indicated below. Which of the functions can be represented by this network?



7. We'd like to consider using a neural net with just one hidden layer, but have it be larger – a hidden layer of size 2. Let's first consider using just two affine functions, with no nonlinearity in between. Which of the functions can be represented by this network?



8. Now we'll add a non-linearity between the two affine layers, to produce the neural network below with a hidden layer of size 2. Which of the functions can be represented by this network?

