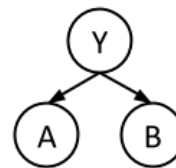


# 1 Naive Bayes

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



- (a) 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	

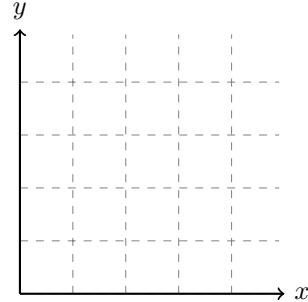
- (b) Consider a new data point ( $A = 1, B = 1$ ). What label would this classifier assign to this sample?
- (c) Let's use Laplace Smoothing to smooth out our distribution. Compute the new distribution for  $P(A|Y)$  given Laplace Smoothing with  $k = 2$ .

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

# 2 Perceptron

You want to predict if movies will be profitable based on their screenplays. You hire two critics A and B to read a script you have and rate it on a scale of 1 to 4. The critics are not perfect; here are five data points including the critics' scores and the performance of the movie:

#	Movie Name	A	B	Profit?
1	Pellet Power	1	1	-
2	Ghosts!	3	2	+
3	Pac is Bac	2	4	+
4	Not a Pizza	3	4	+
5	Endless Maze	2	3	-



- (a) Plot the data above and determine if the points are linearly separable.
- (b) Now you decide to use a perceptron to classify your data. Suppose you directly use the scores given above as features, together with a bias feature. That is  $f_0 = 1$ ,  $f_1 =$  score given by A and  $f_2 =$  score given by B.

Run one pass through the data with the perceptron algorithm, filling out the table below. Go through the data points in order, e.g. using data point #1 at step 1.

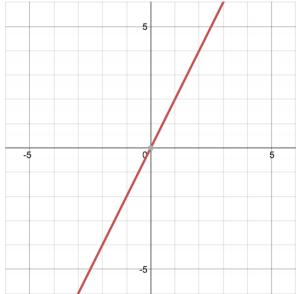
step	Weights	Score	Correct?
1	$[-1, 0, 0]$	$-1 \cdot 1 + 0 \cdot 1 + 0 \cdot 1 = -1$	yes
2			
3			
4			
5			

Final weights:

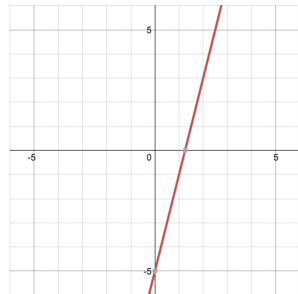
- (c) Have weights been learned that separate the data?
- (d) More generally, irrespective of the training data, you want to know if your features are powerful enough to allow you to handle a range of scenarios. Circle the scenarios for which a perceptron using the features above can indeed perfectly classify movies which are profitable according to the given rules:
  - (a) Your reviewers are awesome: if the total of their scores is more than 5, then the movie will definitely be profitable, and otherwise it won't be.
  - (b) Your reviewers are art critics. Your movie will be profitable if and only if each reviewer gives either a score of 2 or a score of 3.
  - (c) Your reviewers have weird but different tastes. Your movie will be profitable if and only if both reviewers agree.

# 3 Neural Network Representations

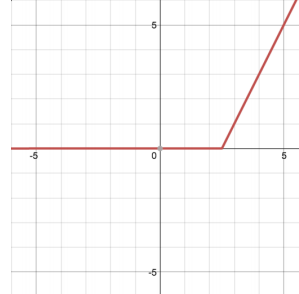
You are given a number of functions (a-h) of a single variable,  $x$ , which are graphed below. The computation graphs on the following pages will start off simple and get more complex, building up to neural networks. For each computation graph, indicate which of the functions below they are able to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



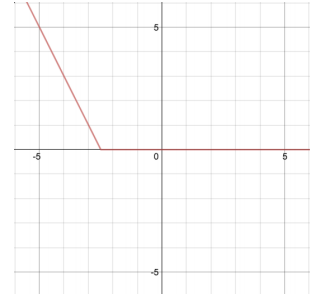
(a)  $2x$



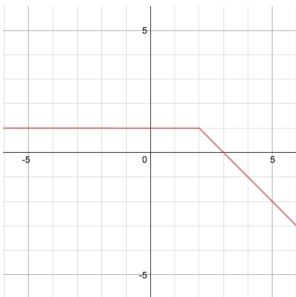
(b)  $4x - 5$



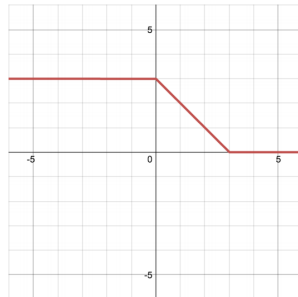
(c)  $\begin{cases} 2x - 5 & x \geq 2.5 \\ 0 & x < 2.5 \end{cases}$



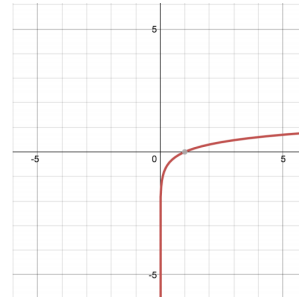
(d)  $\begin{cases} -2x - 5 & x \leq -2.5 \\ 0 & x > -2.5 \end{cases}$



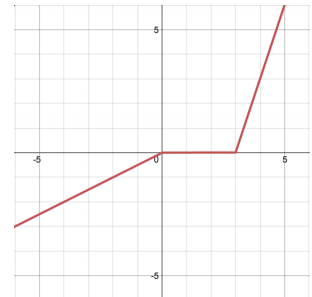
(e)  $\begin{cases} -x + 3 & x \geq 2 \\ 1 & x < 2 \end{cases}$



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



(g)  $\log(x)$

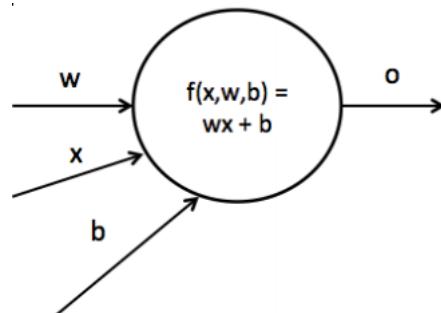
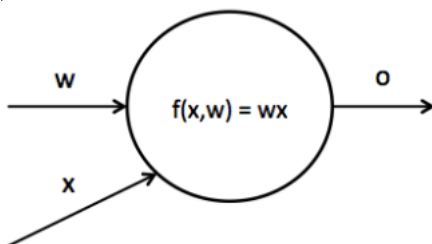


(h)  $\begin{cases} 0.5x & x \leq 0 \\ 0 & 0 < x \leq 3 \\ 3x - 9 & x > 3 \end{cases}$

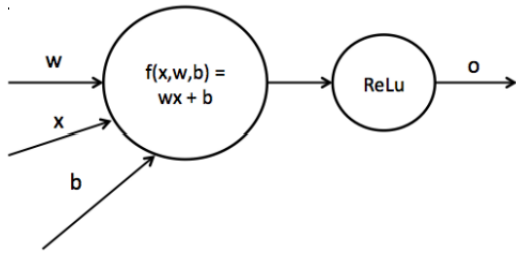
For each of the following computation graphs, determine which functions can be represented by the graph. In parts 1-5, write out the appropriate values of all  $w$ 's and  $b$ 's for each function that can be represented.

1. Linear Transformation

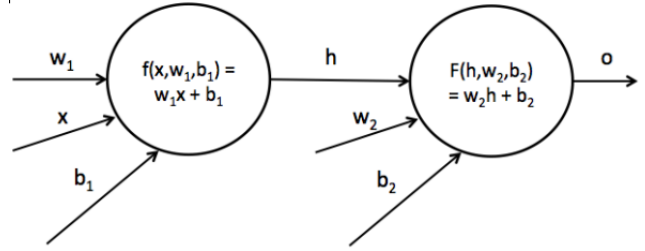
2. Linear plus Bias (aka affine transformation)



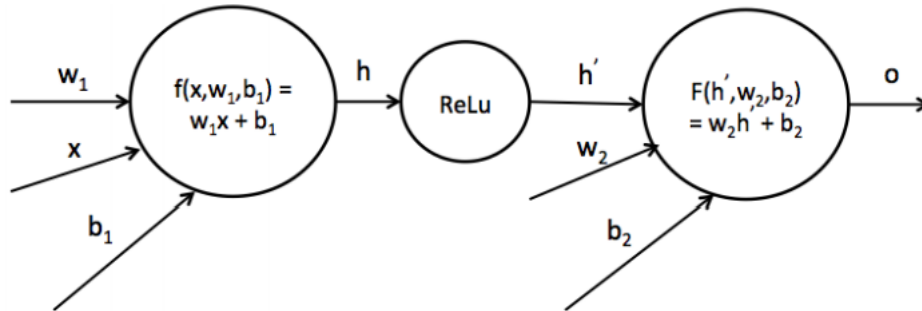
3. Nonlinearity after Linear layer



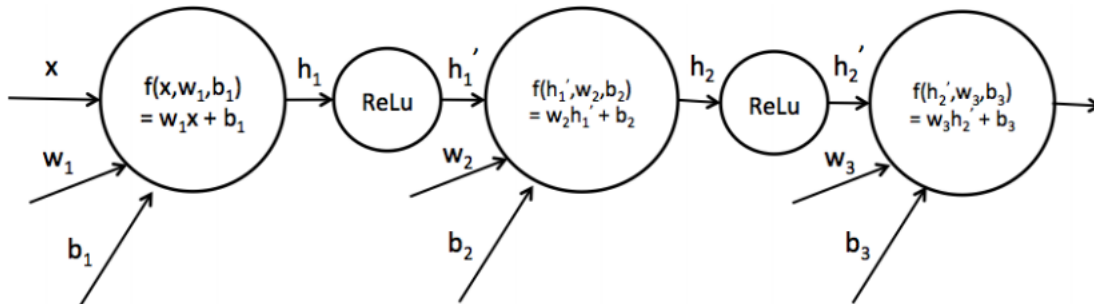
4. Composition of Affine layers



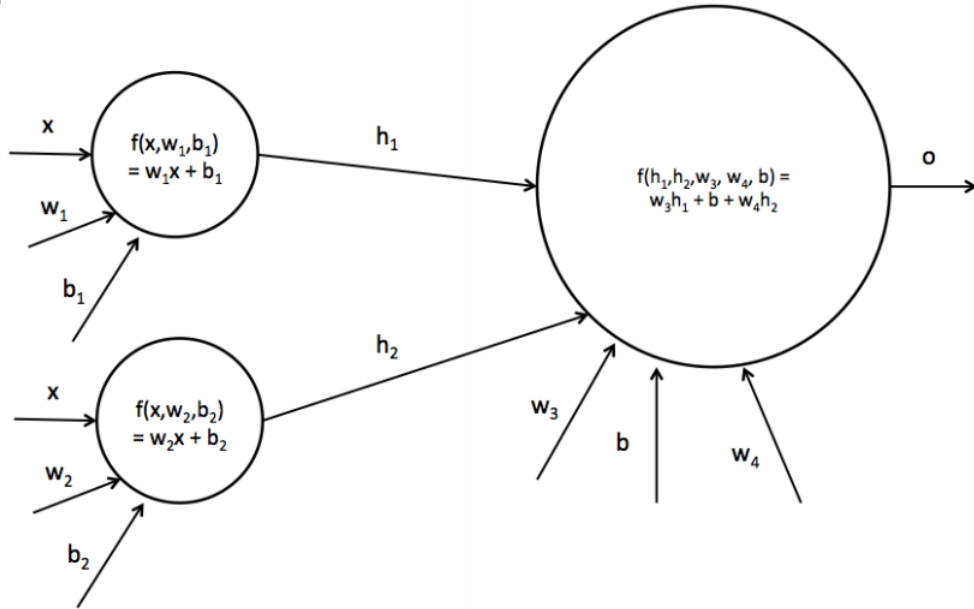
5. Two Affine layers with nonlinearity in between (hidden layer)



6. Add another hidden layer



7. Hidden layer of size 2, no nonlinearities



8. Add nonlinearities between layers

