CSE 5693 Machine Learning HW3
Due 7pm, March 15
Troy Toggweiler

1.a: 4.1

Intersections: x1(-1,0), x2(0,2)
Function: f(x) = (w0 * x0) + (w1 * x1) + (w2*x2)
f(-1) = 0 = (w0*1) + (w1 * -1) + (w2 * 0)
f(0) = 2 = (w0*1) + (w1 * 0) + (w2 * 2)
...Remove zeros
f(-1) = 0 = (w0) + (w1 * -1)
f(0) = 2 = (w0) + (w2 * 2)
...Reduce
w0 = w1
w0 = 4w2
...Substitue
f(0) = 2 = 4w2 + 2w2 => w2 = 1/3
f(0) = 2 = w0 + 2/3 ==> w1 = w0 = 4/3

1.b: 4.2

A & !B

Input Value	Weighted Value: wA=0.5, wB=-0.5	Threshold > 0
A = 1, B = 1	AwA = 0.5, BwB = -0.5 = 0	-1
A = 1, B = 0	AwA = 0.5, BwB = 0 = 0.5	1
A = 0, B = 1	AwA = 0, BwB = -0.5 = -0.5	-1
A = 0, B = 0	AwA = 0, BwB = 0 = 0	-1

A XOR B

Input Value	wA0=0.6, wA1=1.1, wB0=0.6, wB1=1.1	w0=-2 , w1 = 1.1	Threshold > 1
A=1, B=1	A*wA0+B*wB0=1.2 , A*wA1+B*wB1= 2.2	1*w0= -2, 1*w1=1.1 => -0.9	-1
A=1, B=0	A*wA0+B*wB0=0.6 , A*wA1+B*wB1= 1.1	-1*w0= 2, 1*w1=1.1 => 3.1	1
A=0, B=1	A*wA0+B*wB0=0.6 , A*wA1+B*wB1= 1.1	-1*w0= 2, 1*w1=1.1 => 3.1	1
A=0, B 0	A*wA0+B*wB0= 0 , A*wA1+B*wB1= 0	-1*w0= 2, -1*w1=-1.1 => 0.9	-1

1.c: 4.9

No, because the output layer would simply be a weight multiplied by a value, which would not be able to map to required outputs. For example, assuming the best case with a exampleToValue mapping..

Inputs	Hidden values	outputs
1 0 0 0 0 0 0 0	-> 1 ->	1 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0	-> 2 ->	0 1 0 0 0 0 0 0
0 0 1 0 0 0 0 0	-> 3 ->	0 0 1 0 0 0 0 0
0 0 0 1 0 0 0 0	-> 4 ->	0 0 0 1 0 0 0 0
0 0 0 0 1 0 0 0	-> 5 ->	0 0 0 0 1 0 0 0
0 0 0 0 0 1 0 0	-> 6 ->	0 0 0 0 0 1 0 0
0 0 0 0 0 0 1 0	-> 7 ->	0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 1	-> 8 ->	0 0 0 0 0 0 0 1

It is simply not possible to multiplex a single value to eight outputs through simple multiplication and threshold.

1.d:

i. Why does 4 units also work

The nodes in the hidden layer are just some encoding of the inputs. They are multiplexing the inputs across each node. This means that for some given values of in the hidden layer, you can work backwards and get the inputs for them. The more nodes there are the greater the capacity of the multiplex.

What do the hidden values represent

An encoding of the inputs. As discussed above.

Is the magnitude of hidden values significantly different? If so why?

No. due to the sigmoid function they will always be between 0 and 1. So the numbers will naturally gravitate to multiples of 1/n, where N is the number of outputs.

ii. Compare performance of using validation set with noisy iris

Not completed