Problem Set 1:

Josh Adams
jadams334@gatech.edu

A) WE WOULD LIKE TO BUILD A LEARNING ALGORITHM THAT WILL COMPUTE THE PROBABILITY THAT A PARTICULAR KIND OF PERSON HAS CHICKEN POX. SO, GIVEN A SET OF TRAINING DATA WHERE EACH X IS MAPPED TO 1 FOR TRUE OR 0 FOR FALSE:

a.1 Derive the proper error function to use for finding the ML hypothesis using Bayes Rule. You should go through a similar process as the one used to derive least squared error in the lessons.

a.2 Compare and contrast your result to the rule we derived for a deterministic function perturbed by zero-mean gaussian noise. What would a normal neural network using sum of squared errors do with these data? What if the data consisted of x,y pairs where y was an estimate of the probability instead of 0s and 1s?

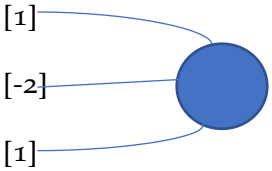
B) DESIGN A TWO-INPUT PERCEPTRON THAT IMPLEMENTS THE BOOLEAN FUNCTION $A \wedge \neg B$. DESIGN A TWO-LAYER NETWORK OF PERCEPTRONS THAT IMPLEMENTS $A \oplus B$ ($\oplus$ IS XOR).

			2	-2	1	
A	B	Y	W_1	W_2	W_0	
0	0	0	2 * 0 + -2 * 0 + 1 * 0 = 0			0
0	1	0	2 * 0 + -2 * 1 + 1 * 0 = -2			0
1	0	1	2 * 1 + -2 * 0 + 1 + 1 = 3			1
1	1	0	2 * 1 + -2 * 1 + 1 * 0 = 0			0

Part 2

1	-2	1	1

W_1 W_2 W_3 Theta



$X_1 \in \{0, 1\}, \quad X_2 \in \{0, 1\}$

Weights
$W_1 = \frac{1}{2}$
$W_2 = -\frac{1}{2}$
$\Theta = \frac{1}{4}$

A	B	(A ∧ ¬B)
F	F	F
F	T	F
T	F	T
T	T	F

$(X_1 * W_1) + (X_2 * W_2) = \Theta$

$X_1 = 0$

$(0 * \frac{1}{2}) + (X_2 * -\frac{1}{2}) = \Theta$

$(X_2 * -\frac{1}{2}) = \frac{1}{4}$

$X_2 = -\frac{0.25}{0.5}$

$X_2 = -0.5 \Rightarrow [0, -0.5]$

$(X_1 * W_1) + (X_2 * W_2) = \Theta$

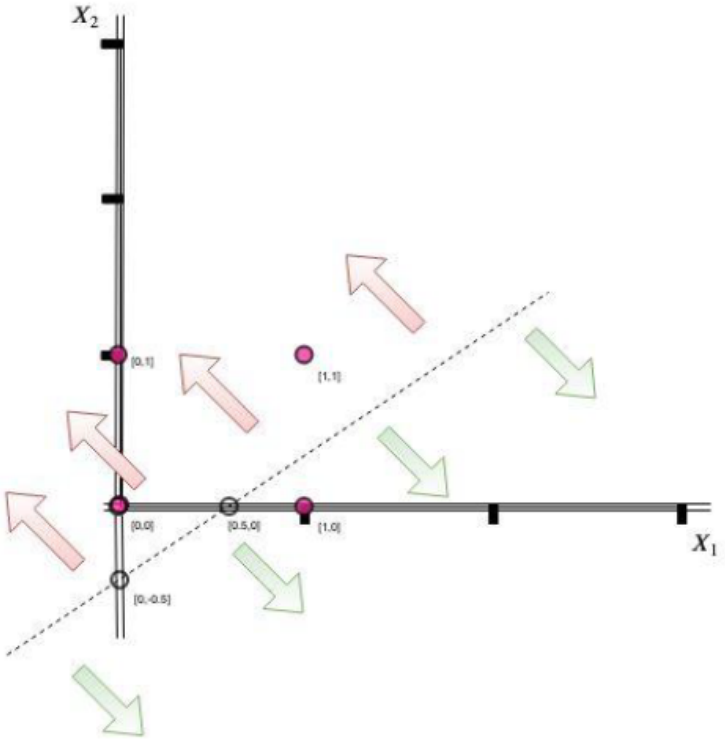
$X_2 = 0$

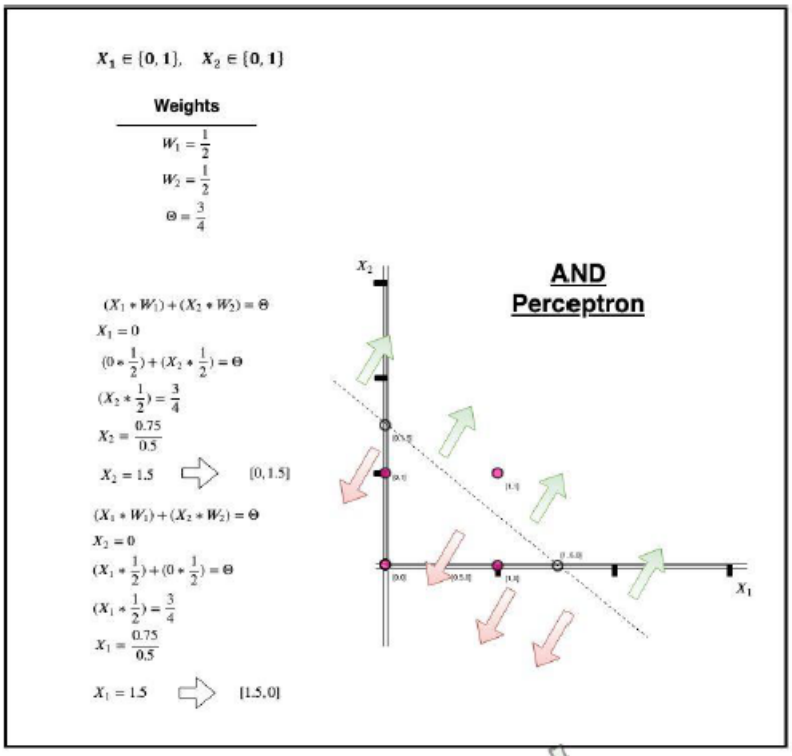
$(X_1 * \frac{1}{2}) + (0 * -\frac{1}{2}) = \Theta$

$(X_1 * \frac{1}{2}) = \frac{1}{4}$

$X_1 = \frac{0.25}{0.5}$

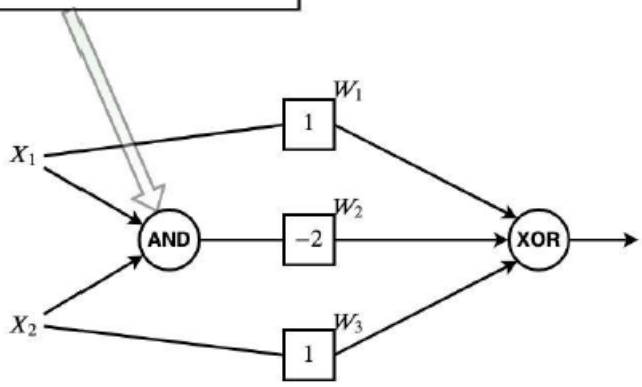
$X_1 = 0.5 \Rightarrow [0.5, 0]$





X_1	X_2	AND	OR	$XOR\ (OR - AND)$
0	0	0	0	0
0	1	0	1	1
1	0	0	1	1
1	1	1	1	0

$Weights$
$W_1 = 1$
$W_2 = -2$
$W_3 = 1$
$\Theta = 1$



C) DERIVE THE PERCEPTRON TRAINING RULE AND GRADIENT DESCENT TRAINING RULE FOR A SINGLE UNIT WITH OUTPUT o, WHERE $o = w0 + w1x1 + w1x\ 2\ 1 + + wnxn + wnx\ 2\ n$. WHAT ARE THE ADVANTAGES OF USING GRADIENT DESCENT TRAINING RULE FOR TRAINING NEURAL NETWORKS OVER THE PERCEPTRON TRAINING RULE?

Im?

D) EXPLAIN HOW YOU CAN USE DECISION TREES TO PERFORM REGRESSION? SHOW THAT WHEN THE ERROR FUNCTION IS SQUARED ERROR, THEN THE EXPECTED VALUE AT ANY LEAF IS THE MEAN. TAKE THE BOSTON HOUSING DATASET (HTTPS://ARCHIVE.ICS.UCI.EDU/ML/DATASETS/HOUSING) AND USE DECISION TREES TO PERFORM REGRESSION.

Decision trees can perform regression as there are classification and regression trees. Regression trees typically use the sum of squared error for breaking up the data. At each node you calculate the average error and make a split based on the best criteria that will reduce the sum of squared error the most. Ultimately this continues until we regress to the mean value.

E) SUGGEST A LAZY VERSION OF THE EAGER DECISION TREE LEARNING ALGORITHM ID3. WHAT ARE THE ADVANTAGES AND DISADVANTAGES OF YOUR LAZY ALGORITHM COMPARED TO THE ORIGINAL EAGER ALGORITHM?

I

F) IMAGINE YOU HAD A LEARNING PROBLEM WITH AN INSTANCE SPACE OF POINTS ON THE PLANE AND A TARGET FUNCTION THAT YOU KNEW TOOK THE FORM OF A LINE ON THE PLANE WHERE ALL POINTS ON ONE SIDE OF THE LINE ARE POSITIVE AND ALL THOSE ON THE OTHER ARE NEGATIVE. IF YOU WERE CONSTRAINED TO ONLY USE DECISION TREE OR NEAREST-NEIGHBOR LEARNING, WHICH WOULD YOU USE? WHY?

Im?

G) GIVE THE VC DIMENSION OF THE FOLLOWING HYPOTHESIS SPACES. BRIEFLY EXPLAIN YOUR ANSWERS SEVEN

g.1 An origin-centered circle (2D)

The VC dimension is 2 because at three points there would be a situation which could not be shattered. Such as 3 points on a line, where the end points are the same label and the middle is different. A circle would not be able to shatter.

g.2 An origin-centered sphere (3D)

The constraint of it being an origin centered sphere makes this have the same VC dimensions as the 2D circle. It would be able to shatter two points but at three it would not be able to

because it is centered. If we were able to move it, then it would be able to shatter and the VC would increase to at least 3.

Answer is VC of 2