6.034 Quiz 3
November 15, 2006

Name	MARVIN MINSKY
EMail	

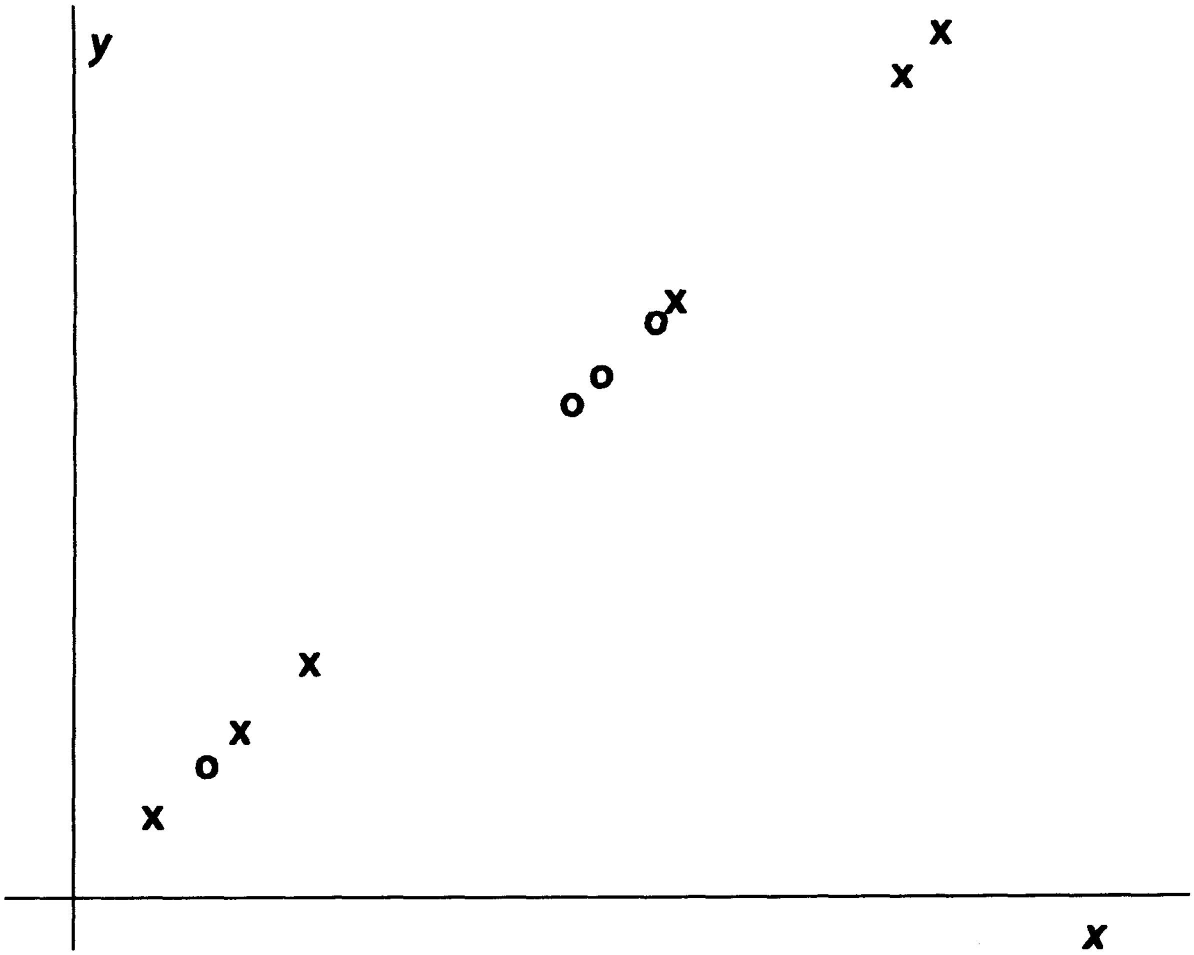
Problem number	Maximum	Score	Grader
1	50	50	
2	50	50	
Total	100	100	

There are **8 pages** in this quiz, including this one.

Problem 1: Nearest Neighbors and Identification Trees (50 points)

Part A: Nearest neighbors (16 points)

David normally thinks of **x**s and **o**s in terms of football plays, but this time, he is thinking about classification using nearest neighbors. He is given the following data:



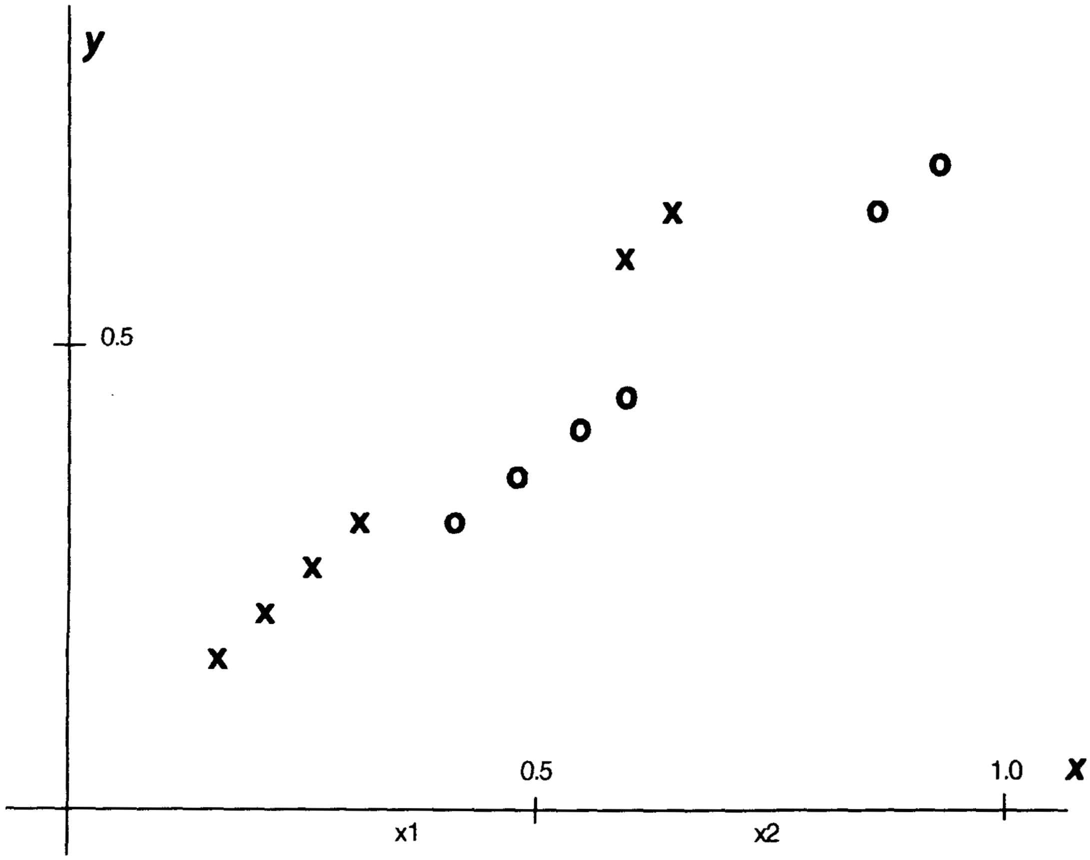
To decide if nearest neighbors is a good idea, David decides to treat of each of the ten points, one by one, as if it were an unknown to be classified rather than a sample. In each case, he asks if the nearest neighbor algorithm would produce the correct classification of the sample from the other nine samples.

Part A1 (8 points)
How many of the 10 samples would David's test correctly identify? 5

Part A2 (8 points)
How many of the 10 samples would David's test correctly identify if he used 3-nearest neighbors rather than just 1-nearest neighbor 8

Part B: Identification trees (34 points)

Laura needs to find a way to identify objects that belong to category x and to category o using the samples provided in the following diagram.



Part B1 (10 points)

Compute the disorder Laura finds for the decision boundary x = x1. You may write your answer in terms of a variable-free expression involving only integers, arithmetic operations, and logarithms.

$$\frac{4}{12}\times 0 + \frac{8}{12}\left(-\frac{2}{8}\log_2\frac{2}{8} - \frac{6}{8}\log_2\frac{6}{8}\right) = \frac{2}{3}H(0.25)$$

Compute the disorder Laura finds for the decision boundary x = x2. You may write your answer in terms of a variable-free expression involving only integers and logarithms.

$$\frac{2}{12}\times 0 + \frac{10}{12}\left(-\frac{4}{10}\log_2\frac{4}{10} - \frac{6}{10}\log_2\frac{6}{10}\right) = \frac{5}{6}H(0.4)$$

$$\frac{5}{6} > \frac{2}{3} \ \& \ H(0.4) > H(0.25)$$

Part B2 (8 points)

On the diagram on the previous page, draw the decision boundaries Laura produces using the identification-tree algorithm. In case two decision boundaries are equally good, use a horizontal rather than a vertical. If there is still a tie, use the smaller threshold. **You will not need to use a calculator to solve this problem.**

Part B3 (8 points)

Patrick writes a program to solve the problem, but he makes a sign error in the method that computs disorder, so it computes the negative of disorder instead of disorder. Write the **disorder** Patrick's program finds for the first decision boundary it draws on the diagram. In this part, we want an **exact numerical answer,** with **no logarithms**.

-1.0

Write the equation for the first **decision boundary** placed by Patrick's program.

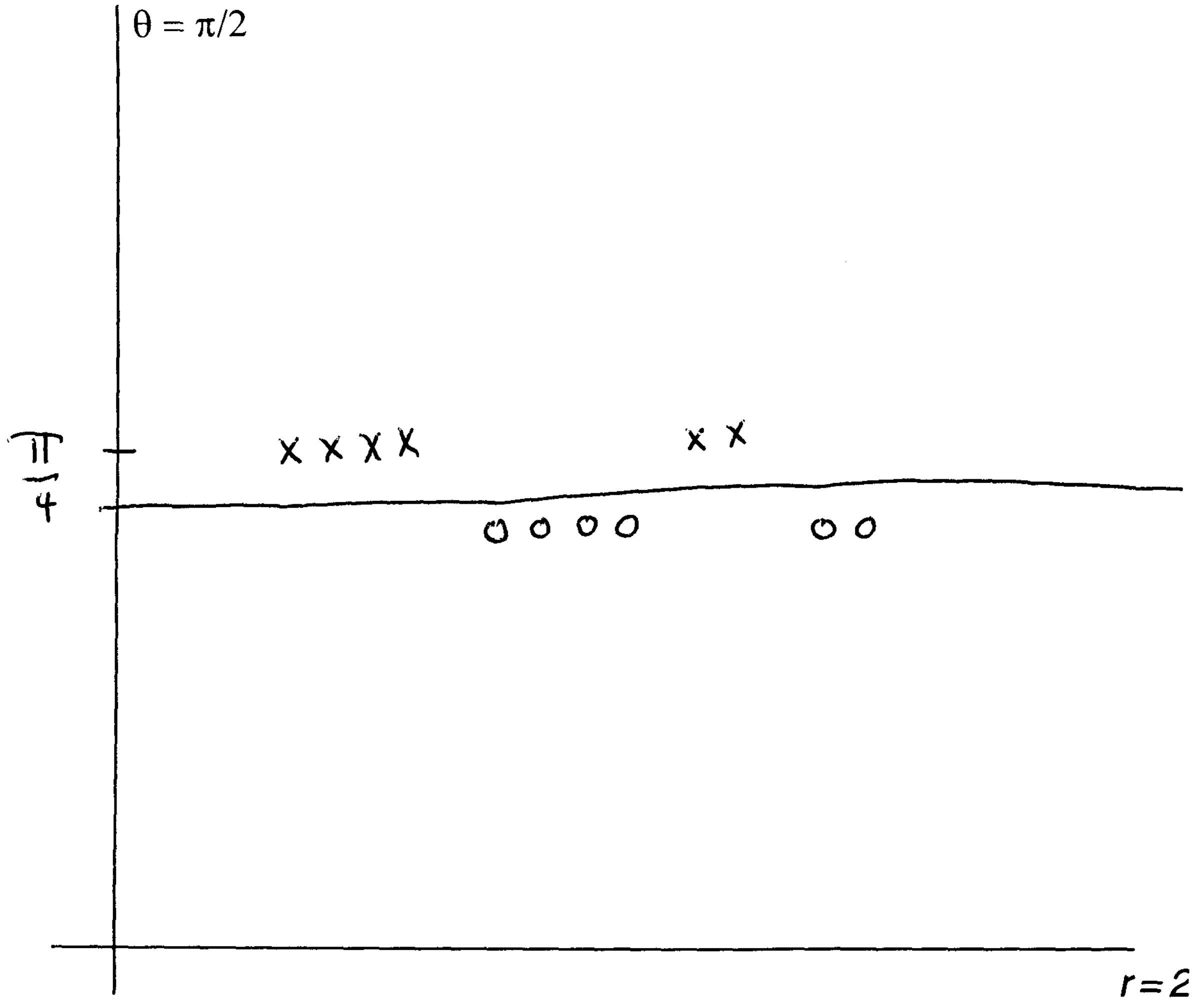
$y = 0.5$

Part B4 (8 points)

Matt suggests that it would be interesting to repeat the algorithm with the data displayed in a polar coordinate system ($\theta = \arctan(y/x)$; $r = \text{sqrt}(x^2 + y^2)$).

Part B4a

Sketch the data on the following coordinate system:

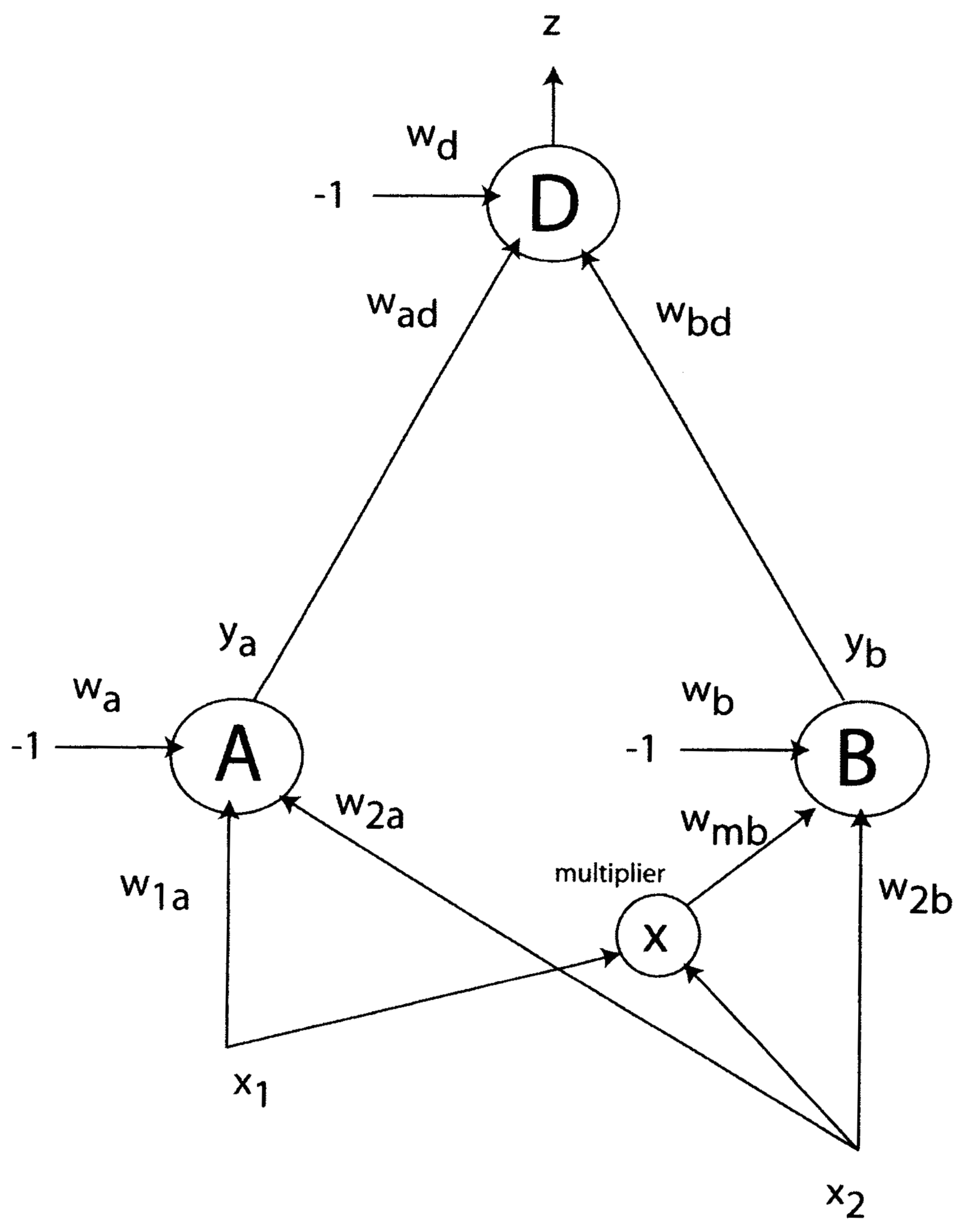


Part4b

Show the decision boundaries he produces by the identification-tree algorithm from these samples.

Problem 2: Neural Nets (50 points)

You now wish to train a neural net using gradient ascent on the standard performance function, $P = -\frac{1}{2}(d-z)^2$. Dissatisfied with linear decision boundaries, you use a net containing a new kind of artificial neuron, B, which is the same as a standard sigmoid unit except that **one input is the product of variables** x_1 and x_2. Both of the other artificial neurons, A and D, are standard sigmoid units.



Part A (17 points)

After running the first example in your training data through this net, you note that the output, z differs from the desired value d. Write an expression for Δw_{bd}, the amount by which w_{bd} is adjusted during backpropagation of the result. The learning rate is r. Write your expression in terms of r, d, z, y_a, y_b, x_1, x_2, and any weights shown on the diagram. Recall that the sigmoid function $o = s(i)$ is $1/(1+e^{-i})$ and its derivative is $o(1-o)$.

Δw_{bd}

$$r\, y_b\, z(1-z)(d-z)$$

mb

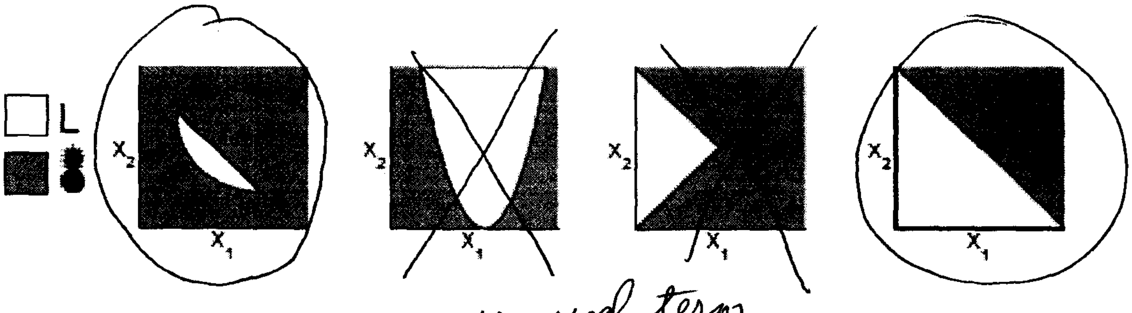
Write an expression for $\Delta w_{\cancel{1b}}$, the amount by which w_{1b} is adjusted during backpropagation of the result. Write your expression in terms of r, d, z, y_a, y_b, x_1, x_2, and any weights shown on the diagram.

Δw_{mb}

$$r\, x_1 x_2\, y_b(1-y_b)\, \omega_{bd}\, z(1-z)(d-z)$$

Part B (12 points)

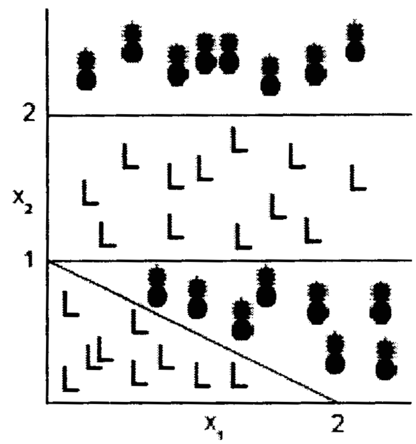
In this part, you are to work with the same neural net as in part A, but **the neurons are not sigmoid units**. Instead, each artificial neuron, A, B, and D is a threshold unit; each outputs 1 if the sum of the weighted inputs is greater than 0, and 0 otherwise. The neural net is to distinguish between pineapples and the letter "L." The output of the net is interpreted as "L" when the output of the net, z, is 1, and "pineapple" otherwise. Which decision boundaries can the net correctly implement? You may assume that the x_1 and x_2 axes have the same scale. **Circle all that apply and cross out those that do not apply.**

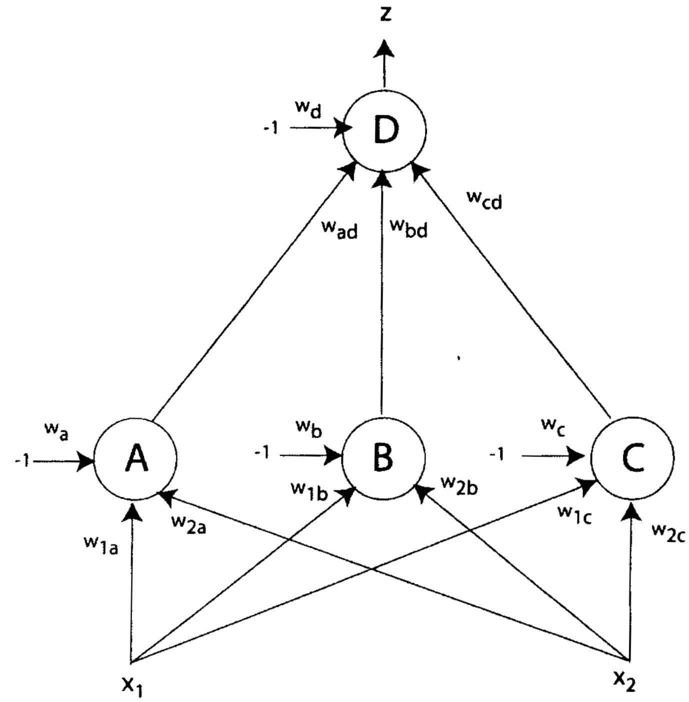


no squared term
$\Rightarrow$ no parabola

Part C (21 points)

In this part, you are to work with a neural net in which **the neurons are not sigmoid units.** Instead, each artificial neuron, A, B, C, and D, is a threshold unit; it outputs 1 if the sum of the weighted inputs is greater than 0, and 0 otherwise. Determine **integer** values for the weights in the network so that it correctly classifies all of the training data. The output of neuron D should be 1 for "L" and 0 for pineapple. Some weights are provided for you.





w_a	-2
W_{1a}	-1
w_{2a}	-2
w_b	1
w_{1b}	0
w_{2b}	1
w_c	-2
w_{1c}	0
w_{2c}	-1
w_d	2
w_{ad}	4
w_{bd}	2
w_{cd}	1

anything ≥ 2

$w_d \parallel w_d - 1$