6.034 Quiz 2
November 15, 2004

Name	
EMail	

Problem number	Maximum	Score	Grader
1	15	15	phw rcb lo sn vm llo lpb js
2	30	30	phw rcb lo sn vm llo lpb js
3	40	40	phw rcb lo sn vm llo lpb js
4	15	15	phw rcb lo sn vm llo lpb js
Total	100	100	

Problem 1: Frames (15 Points)

Select the single **BEST** answer for the following questions. No points can be awarded for multiple selections. No points are deducted for wrong answers.

1. Transition frames make explicit
   1. How state leads to state
   2. How change in state leads to state.
   3. How change in state leads to change in state
   4. How agent classifications shape change in state
   5. How goals and motives shape change in state
   6. All of the above.
   7. None of the above

2. Trajectory frames make explicit
   1. How object classifications govern motion
   2. Plots in the stories that determine cultural bias
   3. How learning progresses with experience
   4. How sentence structure maps into transition-space frames
   5. How object appearance changes with point of view
   6. All of the above
   7. None of the above.

3. The following are readily translated, at least in part, into a transition space:
   1. The car crashed into a wall
   2. The sun rose and then set
   3. The boy picked up a ball
   4. Time stood still
   5. The market fell
   6. None of the above
   7. All of the above

4. Some psychological evidence indicates that knowledge tends to concentrate such that
   1. The most specific classes provide the most knowledge
   2. Intermediate classes provide the most knowledge
   3. The most general classes provide the most knowledge
   4. Knowledge is spread about evenly throughout all classification levels

5. The topological sorting algorithm
   1. Attempts to transform an inheritance hierarchy into a linear precedence ordering
   2. Sometimes cannot find an ordering that satisfies all left-to-right constraints
   3. Sometimes returns a different result from the depth-first up-to-join algorithm
   4. All of the above
   5. None of the above

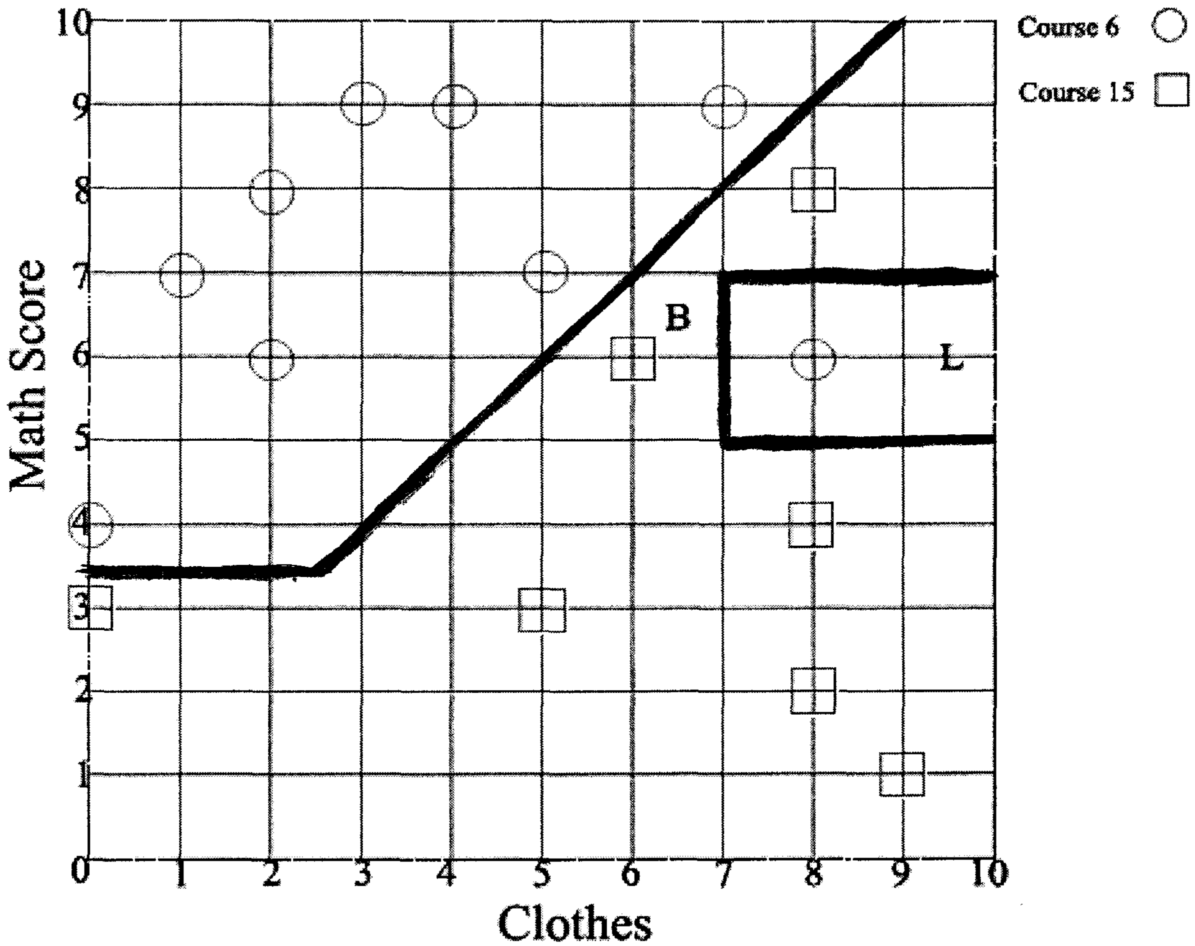
Question 2: Trees and Neighbors (30 points)

As you probably know, Ben Bitdiddle and Louis Reasoner each have little brothers who are freshman in MIT's class of 2008. Ben and Louis have decided to use some machine learning techniques to determine what major would best suite their respective younger siblings.

Part A (10 points)

Ben and Louis collect 16 data points, shown on the graph below (make sure not to miss the two points on the Y axis). The Y axis represents a student's score on the MIT math assessment test (normalized to range from 0 to 10), while the X axis represents how nice a student's clothes are (again normalized to range from 0 to 10). The "B" represents Ben's little brother, while the "L" represents Louis's little brother.

Carefully and precisely draw the decision boundary on the graph above by applying the principle of nearest neighbors (ignore the "B" and "L" points for now).



BF

Part B (2 points)

How is Ben's little brother classified using 1 nearest neighbor?

15

How is Louis's little brother classified using 1 nearest neighbor?

6

Part C (2 points)

"That can't be right" muttered one Course 6 hero to the other. "Must be noisy data…Let's use three nearest neighbors instead."

How is Ben's little brother classified using 3 nearest neighbors?

6

How is Louis's little brother classified using 3 nearest neighbors?

15

Part D (1 point)

Finally, how would each little brother be classified if Ben and Louis were to use 16 nearest neighbors?

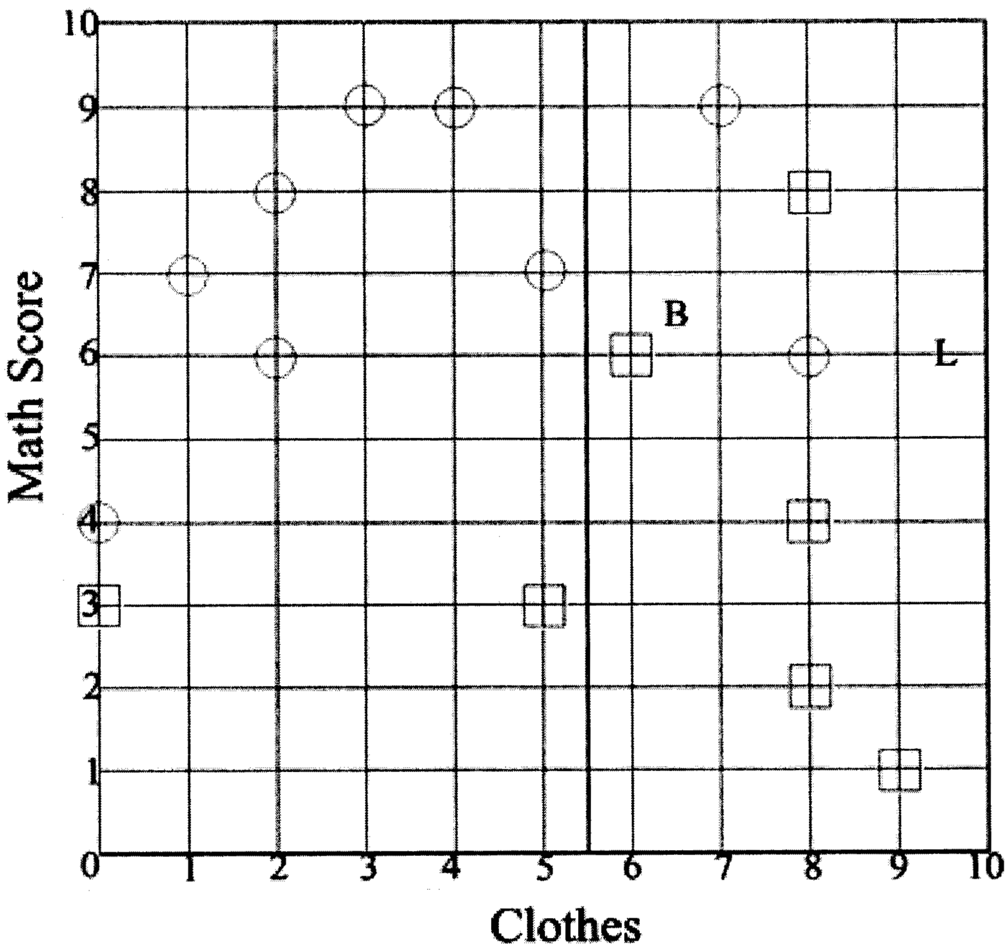
6

Part E (7 points)

"I'm still not convinced this is working. Why don't we give ID trees a shot?" Ben and Louis decide to make their first test be:

Clothes < 5.5

The boundary created by this test is shown on the graph below.



Give the average disorder produced by the test (again ignoring the "B" and "L" points). Your answer may contain logarithms but no variables.

$$\frac{9}{16}\left(-\frac{7}{9}\log_2\frac{7}{9}-\frac{2}{9}\log_2\frac{2}{9}\right)+$$
$$\frac{7}{16}\left(-\frac{2}{7}\log_2\frac{2}{7}-\frac{5}{7}\log_2\frac{5}{7}\right).$$

Part F (8 points)

Further subdivision is needed no matter what the result of the first test. In particular, the samples that fall into the "Yes" category of the first test (those on the left of the vertical line in the graph) still need to be divided further. What test would you use to divide them into groups of minimal average disorder

Math score $\lessgtr 3.5$

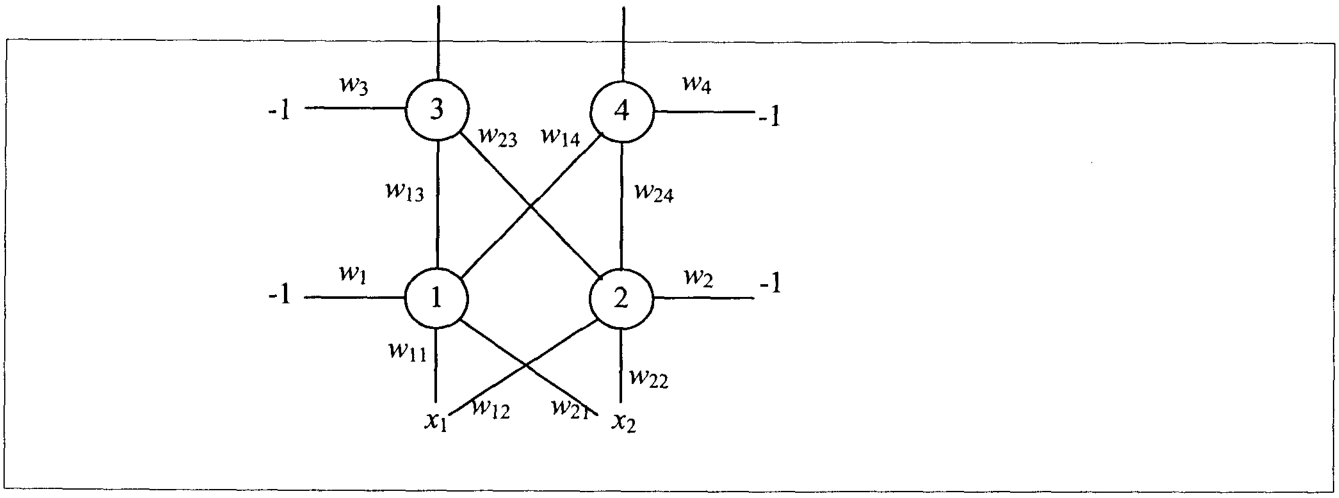
and what would be the resulting average disorder?

0

Question 3: Neural nets (40 points)

The following sigmoid network has 4 units labeled 1, 2, 3, 4. All units are **sigmoid** units, meaning that their output $s(z)$ is computed using the sum of their weighted inputs:

$$s(z) = \frac{1}{1 + e^{-z}}$$



Part A. Forward propagation (8 points)

Using the initial weights provided below, and the input vector $[x_1, x_2] = [-2, 4]$, compute the output at each unit after forward propagation.

Initial weights:

w_1	w_{11}	w_{12}	w_2	w_{21}	w_{22}	w_{13}	w_{14}	w_{23}	w_{24}	w_3	w_4
0.0	2.0	0.5	0.0	1.0	0.25	-8.0	0.0	8.0	0.0	0.0	0.0

A graph of sigmoid values is included for you on the next page.

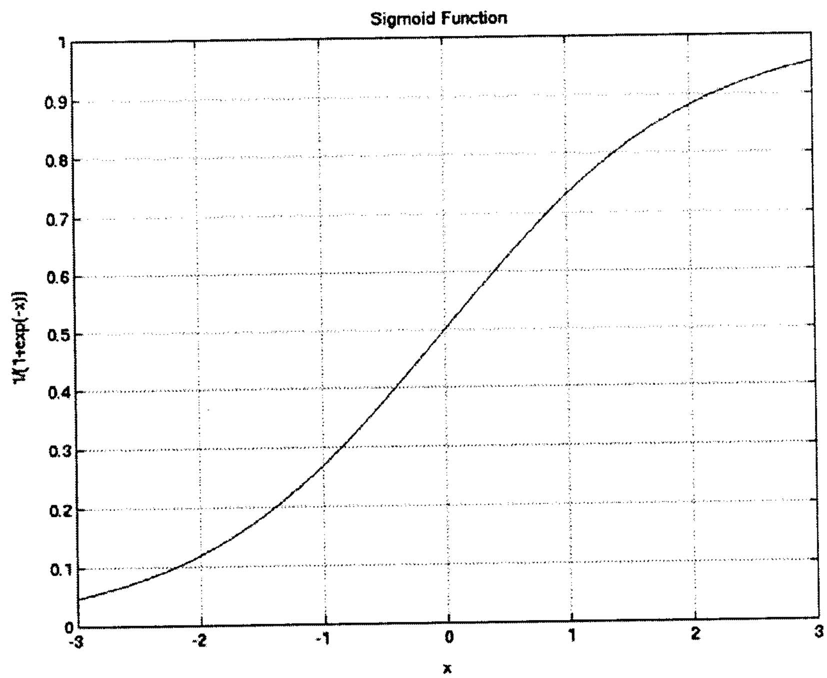
y_1 (output of unit 1) = $s(0(-1) + 2(-2) + 4(1)) = s(0) = .5$

y_2 (output of unit 2) = $s(0(-1) + (.5)(-2) + 4(1)) = s(0) = .5$

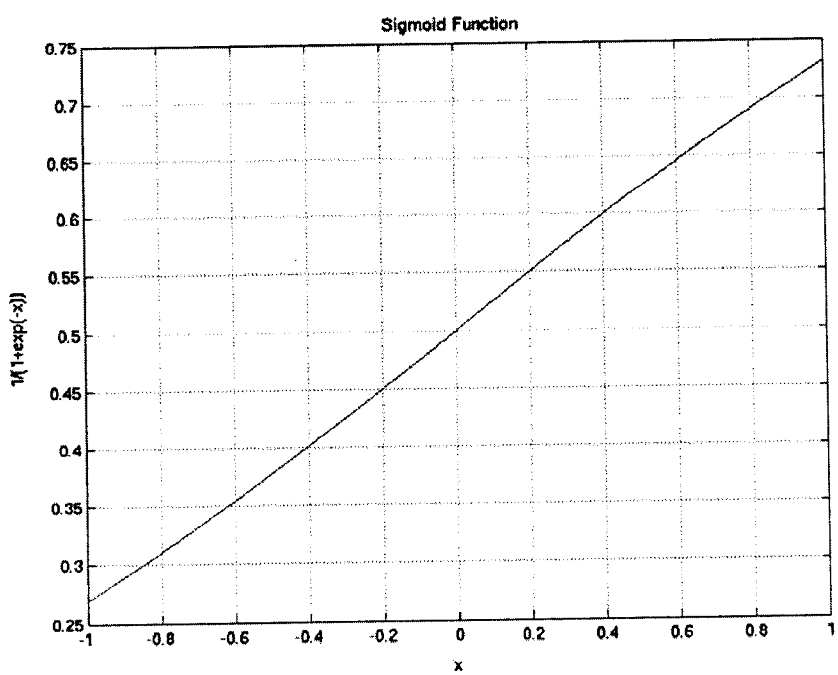
y_3 (output of unit 3) = $s(0(-1) + (-8)(.5) + 8(.5)) = s(0) = .5$

y_4 (output of unit 4) = $s(0(-1) + 0(.5) + 0(.5)) = s(0) = .5$

Wide-angle view



Zoomed view



Part B. Backward propagation (20 points)

For this part of the problem, you may find the notes on the next page helpful; they are drawn from the on-line notes.

Using a learning rate of 4, and desired output of $(y_3^*, y_4^*) = [+1, 0]$, and an error function $E = \frac{1}{2}(y_3 - y_3^*)^2 + \frac{1}{2}(y_4 - y_4^*)^2$ for the input $(x_1, x_2) = [-2, 4]$, backpropagate the network for one iteration, by computing the δ values for nodes 2 and 3, and the new values for the weights w_{23} and w_{22} and filling out the table below.

Express δ_2, and δ_3 in terms of weights, inputs and outputs (no derivatives), and compute their exact numerical values:

	Expression (Derivative free) – 4pts each	*Exact Numerical Value – 1ps each*
δ_2	$y_2(1-y_2)[W_{23}\delta_3 + W_{24}\delta_4] = \frac{1}{2}(1-\frac{1}{2})[8(-\frac{1}{8})] =$	$-\frac{1}{4}$
δ_3	$y_3(1-y_3)(y_3 - y_3^*) = \frac{1}{2}(\frac{1}{2})(\frac{1}{2}-1) =$	$-\frac{1}{8}$

Express the new weights in terms of δs (non-numerical) and numbers.

	Expression using δs – 4pts each	*Exact Numerical Value – 1ps each*
w_{23}	$W_{23} - R \cdot y_2 \cdot \delta_3 = 8 - 4(\frac{1}{2})(-\frac{1}{8}) =$	8.25
w_{22}	$W_{22} - R \cdot X_2 \cdot \delta_2 = \frac{1}{4} - 4(4)(-\frac{1}{4}) =$	4.25

ELF

Backprop notes

$$E = \frac{1}{2}\sum_k (o_k - d_k)^2$$

$$w_{i\to j} = w_{i\to j} - \Delta w_{i\to j}$$

$$\Delta w_{i\to j} = R \times o_{l_i} \times \delta_{r_j}$$

where R is a rate constant and the δs are computed with the following formulas:

$$\delta_k = o_k(1 - o_k) \times (o_k - d_k)$$

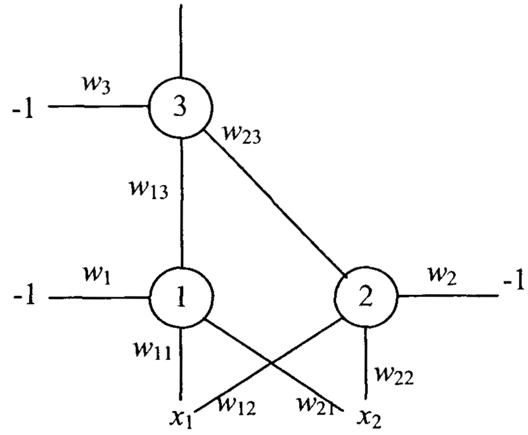
$$\delta_{l_i} = o_{l_i}(1 - o_{l_i}) \times \sum_j w_{i\to j} \times \delta_{r_j}$$

where

- o_k is output k of the output layer
- d_k is the desired output k of the output layer
- δ_k is a delta associated with the output layer
- o_{l_i} is output i of left layer in a left-right pair
- δ_{l_i} is a delta associated with the layer l
- δ_{r_j} is a delta associated with the adjacent layer to the right, layer r

Part C (5 points)

Consider the following network:



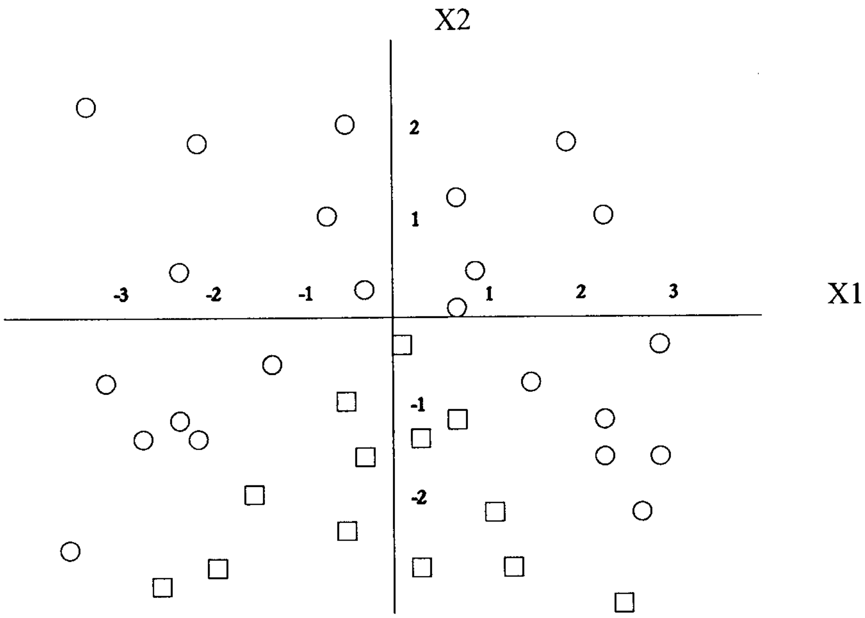
All units have step function thresholds, not sigmoids.

That is, output = 1 for input greater than 0; output = 0 for input equal to or less than zero.

There are five versions corresponding to five sets of weights:

Network	w_1	w_{11}	w_{12}	w_2	w_{21}	w_{22}	w_{13}	w_{23}	w_3
A	0	-1	1	0	1	1	2	2	1
B	0	-1	1	0	1	1	1	1	2
C	0	-1	1	0	1	1	1	1	0
D	0	-2	2	0	2	2	2	2	1
E	0	-2	2	0	2	2	2	2	4

You are to determine which set of weights in the table do well, given that you want unit $y_3 = 1$ for the circular points and $y_3 = 0$ for the square points:



Indicate which of the networks you expect to give nearly zero error on the given data set by circling one or more of the following:

A	B	C	D	E

Part D (4 Points)

Indicate whether the following statement is true or false by writing "true" or "false" in the box, **and** explain why in a few words. Our solution will have 10 words or less.

For a fixed network architecture, the backpropagation algorithm is guaranteed to find the best possible set of weights, given sufficient training data.

FALSE. BACKPROP IS SUSCEPTIBLE TO BEING TRAPPED IN LOCAL MINIMA.

Part E (3 Points)

Indicate whether the following statement is true or false by writing "true" or "false" in the box, **and** explain why in a few words. Our solution will have 10 words or less.

You can always improve performance on both your training and test sets by adding more hidden layers to your network.

FALSE. OVERFITTING ON TRAINING SET MAY LEAD TO POORER PERFORMANCE ON TEST SET.

This page is a spacer. Go to next page.

Problem 4: Support Vector Machines (15 Points)

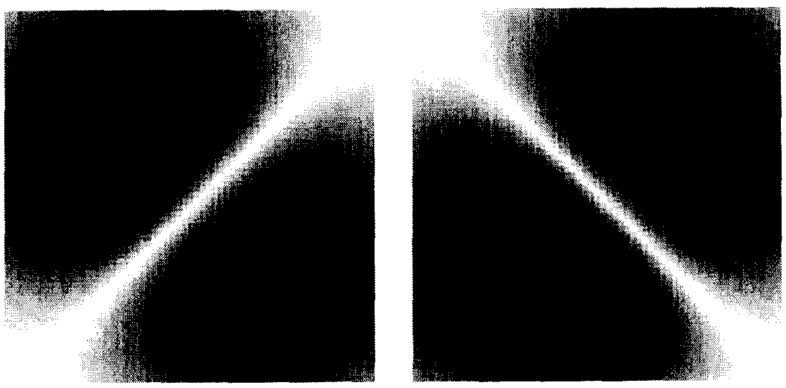
The following diagrams illustrate the behavior of support vector machines that have been trained to separate pluses (+) from minuses (-). The origin is at the lower left corner in all diagrams.

Exactly the same kernel function was used to produce both left and right pictures in each pair, and that kernel has one of two forms:

$$e^{-\frac{\|x_1-x_2\|^2}{\sigma}} \quad \textbf{or} \quad (x_1 \bullet x_2)^n$$

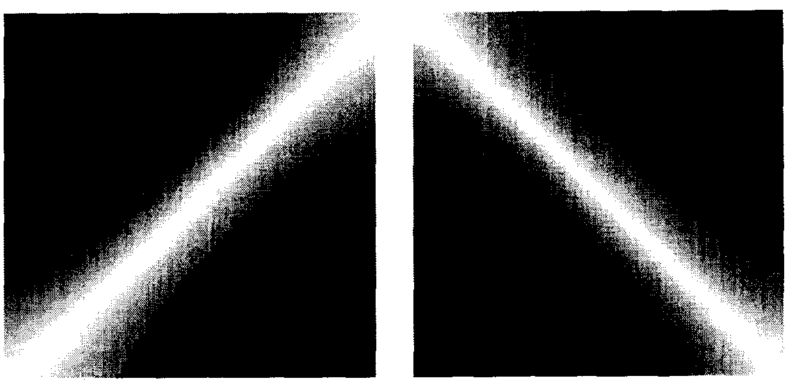
In some diagrams, red vectors are marked because the machine was unable to handle all the samples; it gave up on noting that one or more weights seemed to be climbing toward infinity.

For each pair, circle the kernel that was used to generate both diagrams of each pair.



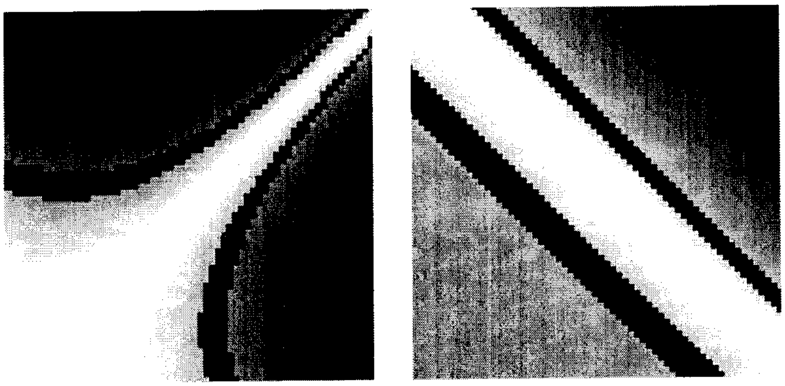
Part 1

Radial basis function, sigma 0.12
Radial basis function, sigma 0.50
Dot product, n = 1
Dot product, n = 3



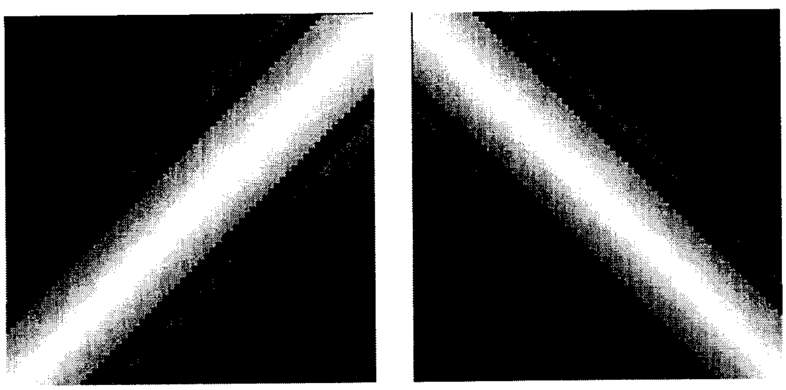
Part 2

Radial basis function, sigma 0.12
Radial basis function, sigma 0.50
Dot product, n = 1
Dot product, n = 3



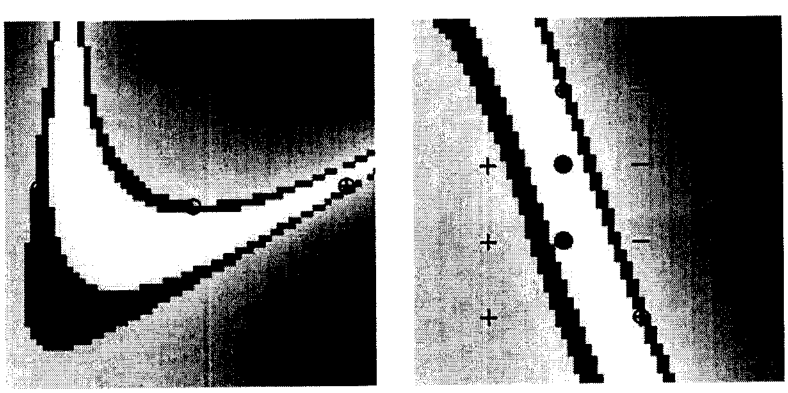
Part 3

Radial basis function, sigma 0.12
Radial basis function, sigma 0.50
Dot product, n = 1
Dot product, n = 3



Part 4

Radial basis function, sigma 0.12
Radial basis function, sigma 0.50
Dot product, n = 1
Dot product, n = 3



Part 5

Radial basis function, sigma 0.12
Radial basis function, sigma 0.50
Dot product, n = 1
Dot product, n = 3