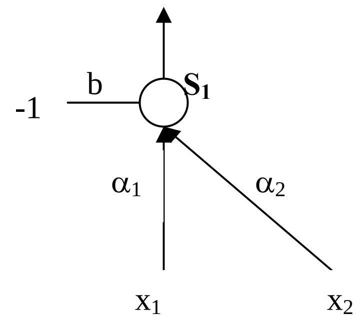
Problem 1:

In this problem you are asked to build a neural network step by step, and then reason about the properties of this network. The problem has many questions but the amount of work required is much less than it may seem at first glance. We first guide you in constructing a particular type of neural network. Then, we ask you to reason about the network you just built.

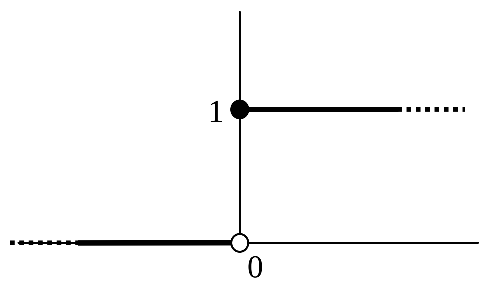
Building one Neural Network Unit (5 parts)



Consider the network unit shown on the right

At first, we are interested in building a network using a perceptron type step function as the threshold unit rather than a sigmoid functions. Formally, the step function threshold unit is defined as follows:

$$S(input) = \begin{cases} 0 & if\ input < 0 \\ 1 & if\ input \geq 0 \end{cases}$$

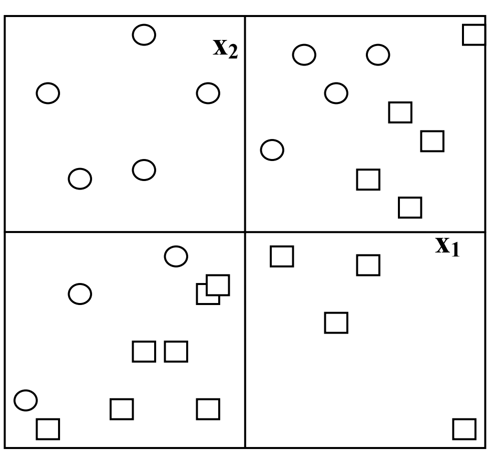


The step function is shown in the figure to the right. **Note the convention adopted for the output when input=0, which will matter later on.** With this definition, the output of the unit in the figure can be written as

$$\mathbf{y_1 = S_1(\alpha_1 x_1 + \alpha_2 x_2 - b).}$$

Where S_1 is a step threshold function.

Consider the following data, in which **squares** stand for class y=0 (**□, y=0)**, and **circles** stand for class y=1**, (○,y=1))**:



1. Draw the decision boundary

2. Write the equation for the decision boundary

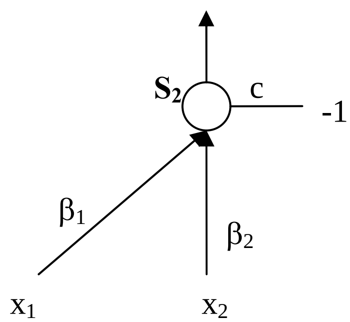
Implementing the decision boundary for the unit above given this sample data implies finding a relationship between the alphas α.

3. What is the relationship between the alphas α?

4. What should the offset (b) be.

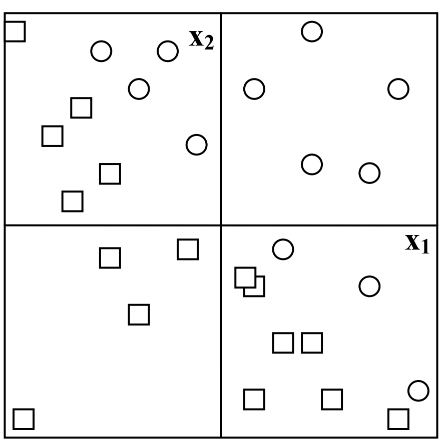
5. Is there any additional constraint on α_1 needed to ensure that the network produces the right answer on each side of the boundary? (Yes/No, if yes, write it down)

Building another Unit (5 parts)



Consider now a second Neural Network unit, like the one on the right. You should use the same step threshold function as in part 1.

Consider the following data:



6. Draw the decision boundary

7. Write the equation for the decision boundary

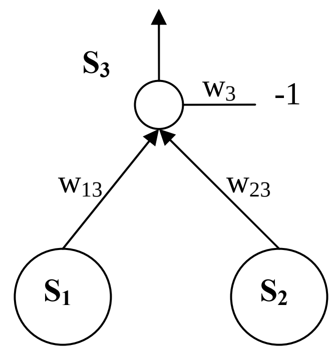
8. What is the relationship between the betas β.?

9. What should the offset (c) be?

10. Is there any additional constraint on β_1. (Yes/No, if yes, write it down)

Building the final unit

We want to construct a third unit that takes as input the outputs of S1 and S2.:

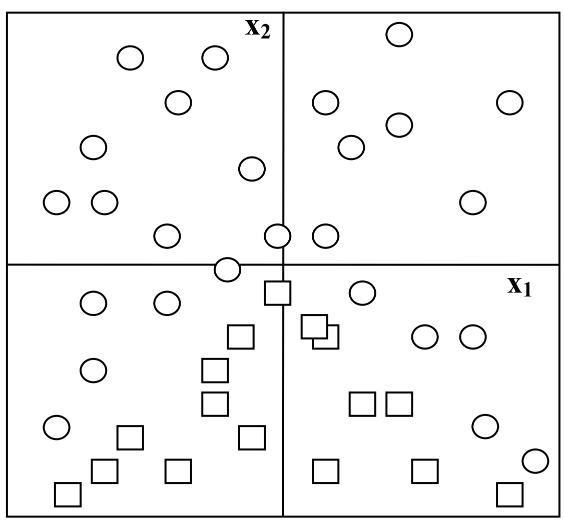


11. Complete the expression for output of the third threshold unit in terms of:

- x_1 and x_2,
- the functions S_1 and S_2,
- and your knowledge about the αs and the βs.

S_3(**)**

For the rest of the problem we will consider the following sample Data.



This data may look familiar, but as a wise Spanish proverb says: "At night, all cats look brown," which implies that you should wait until light shines on the problem before jumping to conclusions.

So are the weights constrained? (4 parts)

Consider the three data points

X_1	x_2
0	-2
3	0
-2	1

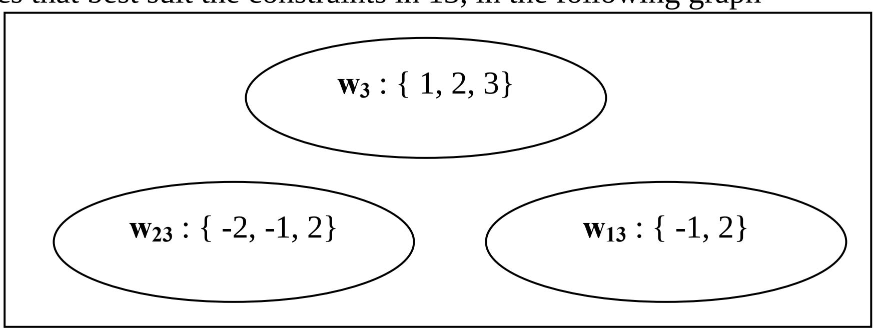
12. Looking at the formula you obtained in part 11, use the table above to determine three constraints on the weights for the final unit (w_{13}, w_{23}, w_3)

c1	**c2**	**c3**

Consider now the following variables and domains:

Variable	Domain
w_{13}	{-1,2}
w_{23}	{-2,-1,2}
w_3	{1,2,3}

13. Draw the arcs that best suit the constraints in 13, in the following graph



14. Run ***constraint propagation*** (a.k.a. ***arc consistency***) on the above graph, and fill the table below as you proceed (if some arc does not exist in your graph, just **cross** it out) :

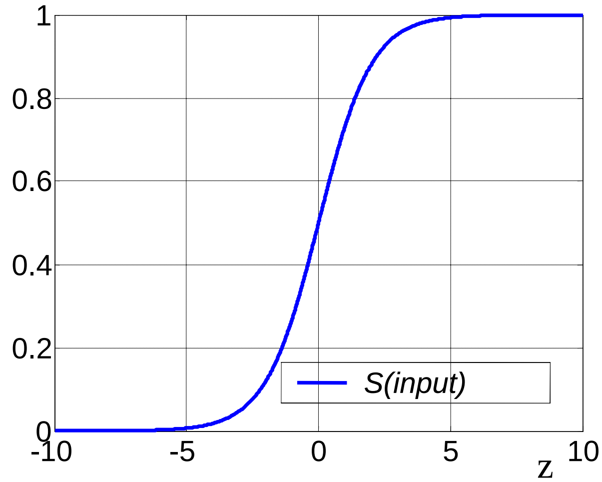
Arc	Resulting domain
$w_3 \rightarrow w_{23}$	$w_3 \in \{ \qquad \}$
$w_{13} \rightarrow w_3$	$w_{13} \in \{ \qquad \}$
$w_{23} \rightarrow w_{13}$	$w_{23} \in \{ \qquad \}$
$w_{13} \rightarrow w_{23}$	$w_{13} \in \{ \qquad \}$
$w_3 \rightarrow w_{13}$	$w_3 \in \{ \qquad \}$
Cassini → **Titan**	**Huygens** ∈ {crashed, landed}
$w_{23} \rightarrow w_3$	$w_{23} \in \{ \qquad \}$

15. What are the possible solution(s) you get?

A sigmoid neural network (5 parts)

Now asume that we want to use sigmoid units instead of step units. Recall that a sigmoid unit would look like:

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



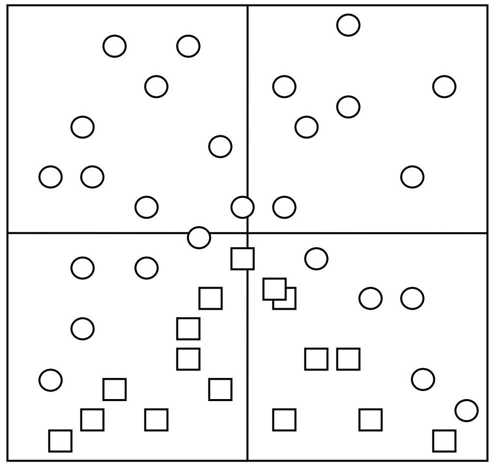
16. Show the new output of $\mathbf{s_3}$ in terms of $\mathbf{x_1}$ and $\mathbf{x_2}$. **Note that s1 and s2 must not appear in your solution.** Look back at question 11 and replace the step functions that appeared there by the sigmoid expression. Set the αs and the βs to +1 or -1 in accordance with your answers to part 5 and part 10. Simplify the result as much as you can.

s$_3$(**)**

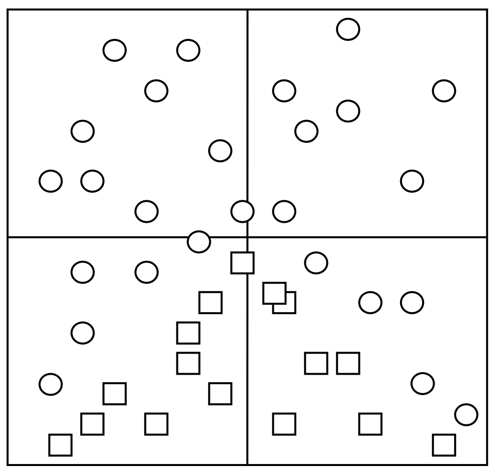
17. Recall that the output of a sigmoid function is continuous. However, for classification, we use the following relation:

$$class = \begin{cases} 0 & \text{if } s(input) < 1/2 \\ 1 & \text{if } s(input) \geq 1/2 \end{cases}$$

Draw *approximately* the decision boundary of the **sigmoid network** if the weights in the final unit were: $\mathbf{w_{13}}$=2, $\mathbf{w_{23}}$=2, $\mathbf{w_3}$=2. and the αs and βs are +1 or -1, as determined in part 16.



18. Do the same assuming the weights were $\mathbf{w_{13}}$=2, $\mathbf{w_{23}}$=2, $\mathbf{w_3}$=1 using the same αs and βs as before.



19. Comment on the most salient difference, if any, between the decision boundaries produced by a sigmoid unit and a step threshold unit.

20. What would you change, if anything, in the sigmoid function to arbitrarily approximate the decision boundary when using the step function?

Problem2:

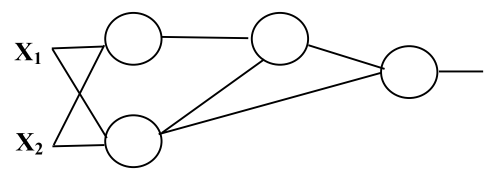
Part A

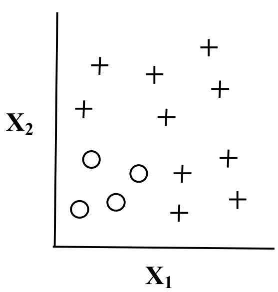
For each of the following data sets, draw the minimum number of decision boundaries that would completely classify the data using a perceptron network.

Part B

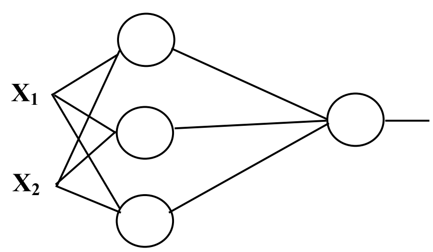
Recall that the output of a perceptron is 0 or 1. For each of the three following data sets, select the perceptron network with the fewest nodes that will separate the classes, and write the corresponding letter in the box. ***You can use the same network more than once.***

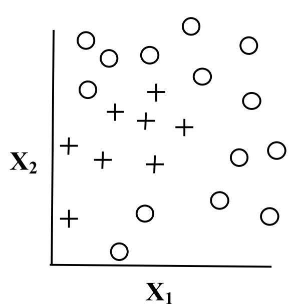
A



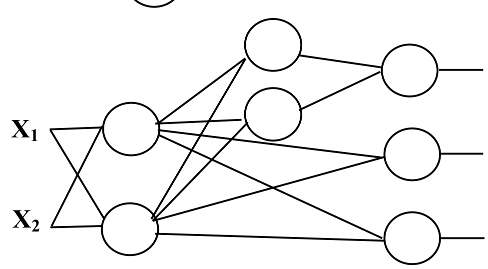


B

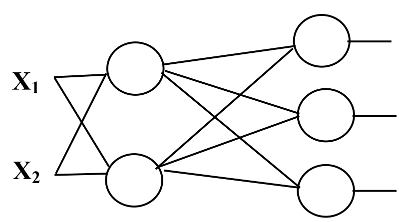


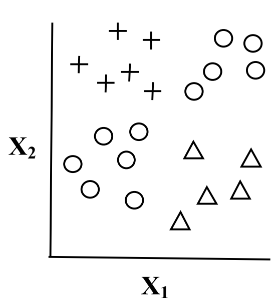


C

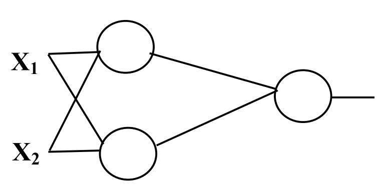


D





E



Problem 3:

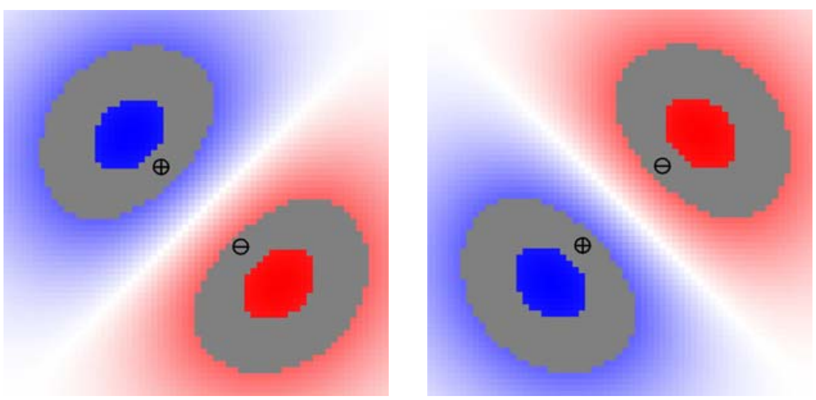
The following diagrams illustrate the behavior of support vector machines that have been trained to separate pluses (+) from minuses (-). The origin is at the lower left corner in all diagrams.

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:

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

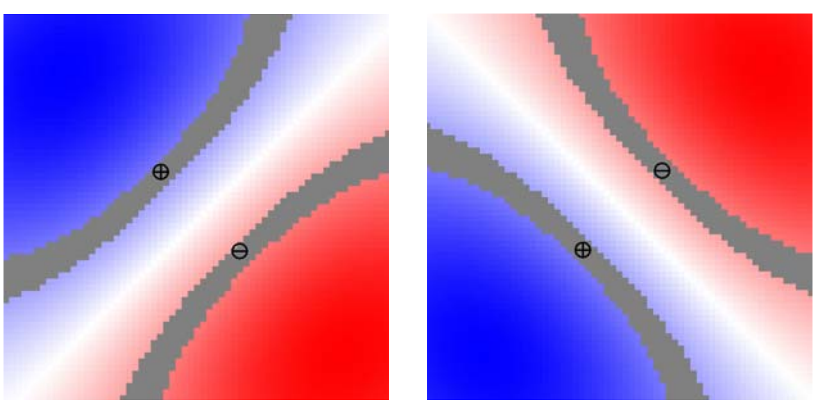
In some diagrams, red vectors are marked because the machine was unable to handle all the samples; it gave up on noting that one or more weights seemed to be climbing toward infinity.

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



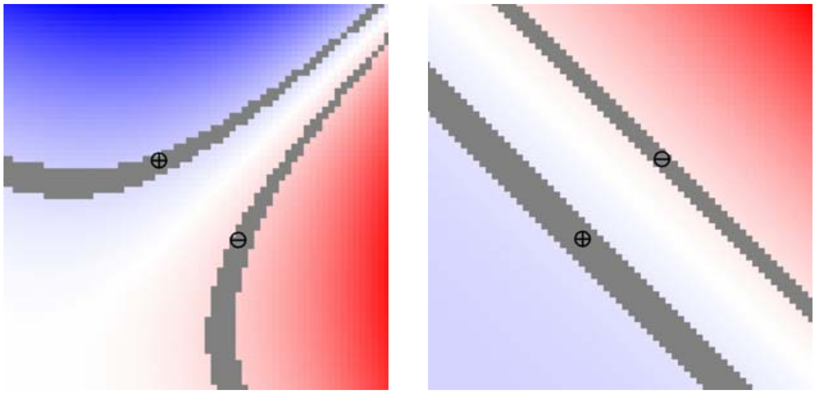
Part 1

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



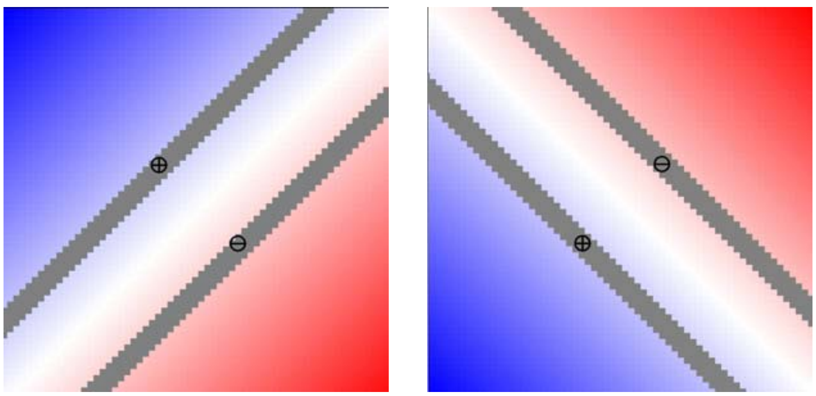
Part 2

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



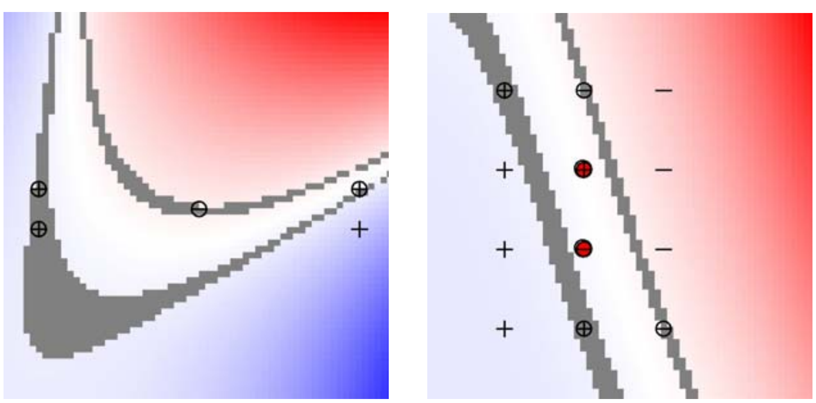
Part 3

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

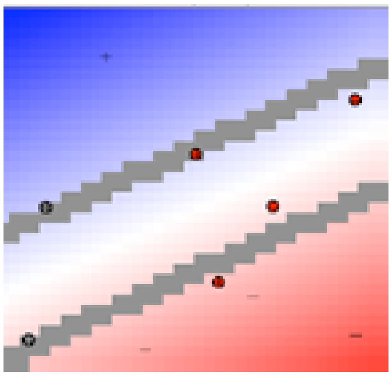
Radial basis function, sigma 0.12
Radial basis function, sigma 0.50
Dot product, n = 1
Dot product, n = 3

Problem 4:

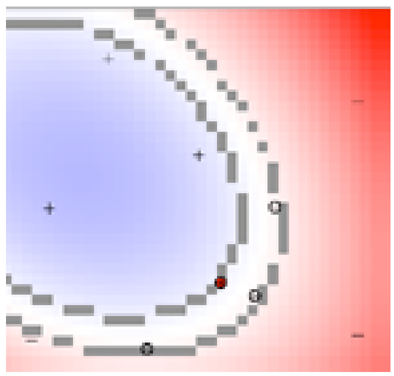
The following diagrams represent graphs of support vector machines trained to separate pluses (+) from minuses (-) for the same data set. The origin is at the lower left corner in all diagrams. Which represents the best classifier for the training data? ***See the separate color sheet for a clearer view of these diagrams.***

Indicate your choice here:

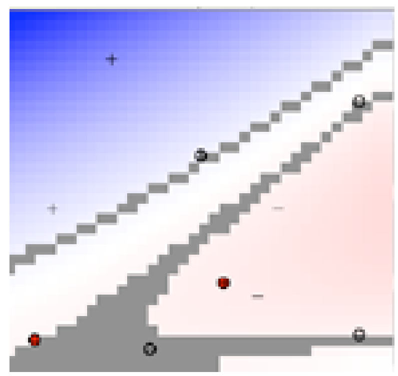
A.



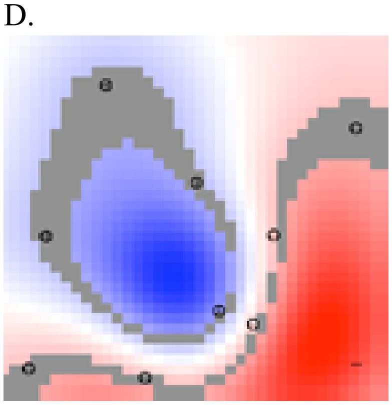
B.



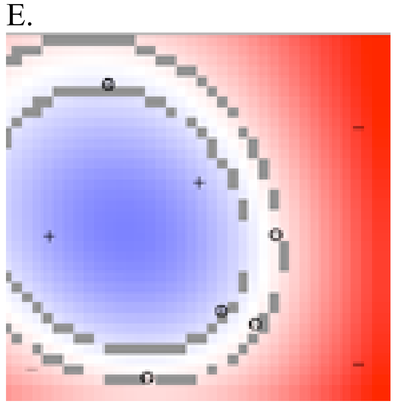
C.



D.



E.



Part B:

Match the diagrams in Part 1 with the following kernels:

Radial basis function, sigma .08

Radial basis function, sigma .5

Radial basis function, sigma 2.0

Linear

Second order polynomial

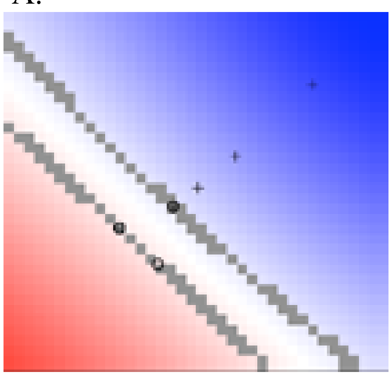
Part C:

Order the following diagrams from ***smallest*** support vector weights to ***largest*** support vector weights, assuming all diagrams are produced by the same mechanism using a linear kernel (that is, there is no transformation from the dot-product space).

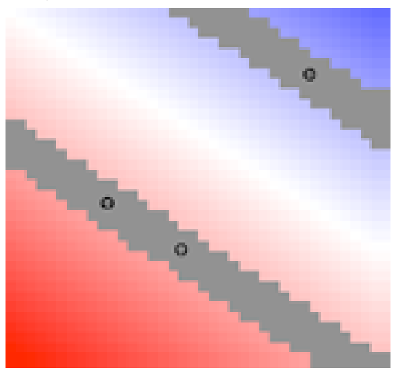
The origin is at the lower left corner in all diagrams. Support vector weights are also referred to as □$_i$ values or LaGrangian multipliers. ***See the separate color sheet for a clearer view of these diagrams.***

Smallest ***Medium*** ***Largest***

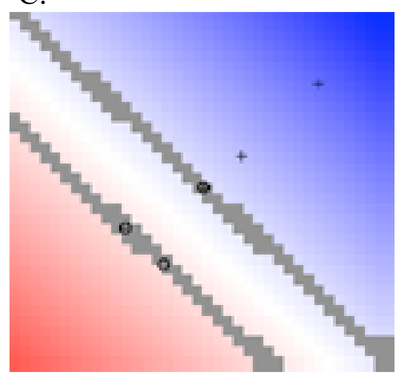
A.



B.



C.



Part D

Suppose a support vector machine for separating pluses from minuses finds a plus support vector at the point $\mathbf{x_1}$= (1, 0), a minus support vector at $\mathbf{x_2}$= (0, 1).

You are to determine values for the classification vector ***w*** and the threshold value b. Your expression for ***w*** may contain $\mathbf{x_1}$ and $\mathbf{x_2}$ because those are vectors with known components, but you are not to include any □$_i$ or y_i. Hint: think about the values produced by the decision rule for the support vectors, $\mathbf{x_1}$ and $\mathbf{x_2}$.

w

b