Final Exam 2002 Problem 6: Support Vector Machines (14 Points)

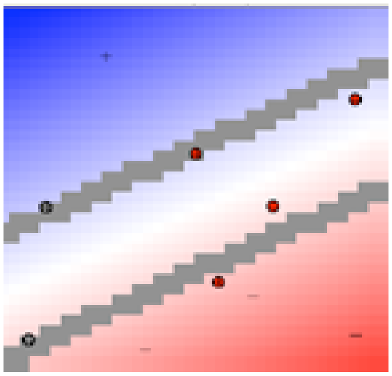
Part A: (2 Points)

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.***

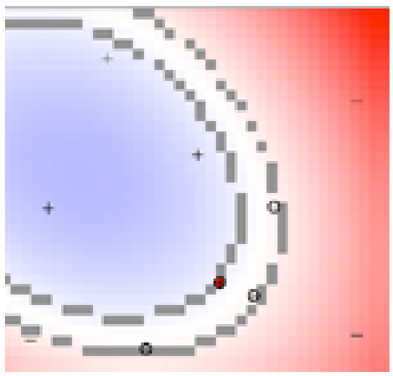
Indicate your choice here: E

E gives proper values to all points. So does D, but at the risk of overfitting.

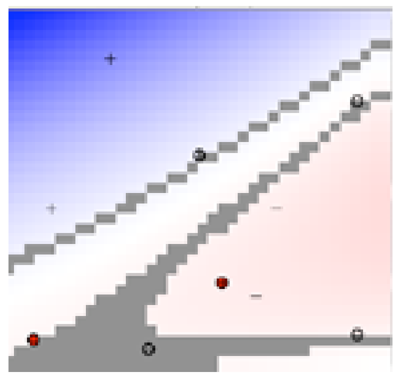
A.



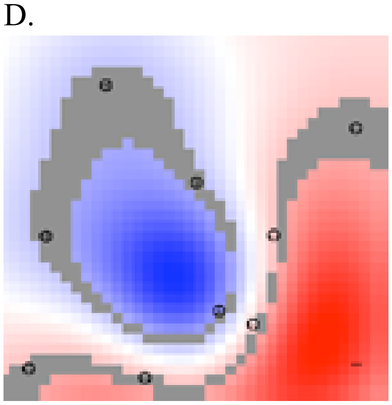
B.



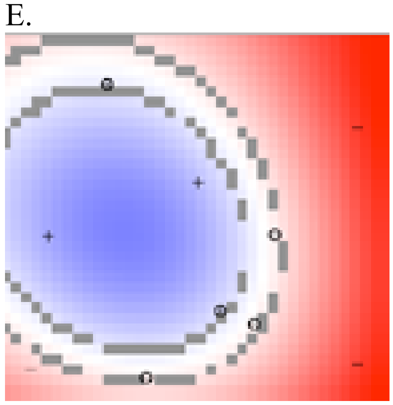
C.



D.



E.



Part B: (5 Points)

Match the diagrams in Part 1 with the following kernels:

Radial basis function, sigma .08	D
Radial basis function, sigma .5	E
Radial basis function, sigma 2.0	B
Linear	A
Second order polynomial	C

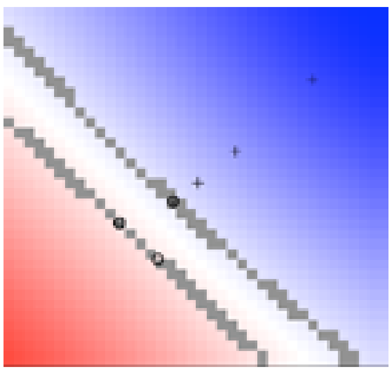
Part C: (3 Points)

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).

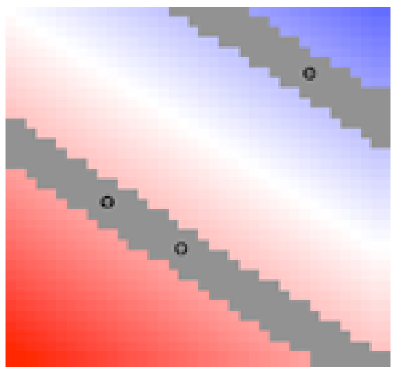
The origin is at the lower left corner in all diagrams. Support vector weights are also referred to as α_i values or LaGrangian multipliers. ***See the separate color sheet for a clearer view of these diagrams.***

Smallest	***Medium***	***Largest***
B	C	A

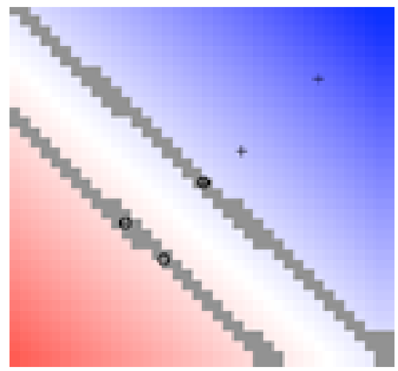
A.



B.



C.



Part D (4 Points)

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

$\boldsymbol{w}$ **$\mathbf{x_1} - \mathbf{x_2}$ = [1 –1]**

b 0

The decision boundary goes through the origin, so b = 0.

We know $\boldsymbol{w} = (a_1\,\mathbf{x_1} - a_2\,\mathbf{x_2})$, and $\mathbf{x_1}$ is a support vector so, $\boldsymbol{w} \,.\, \mathbf{x_1} = 1$. Substituting for $\boldsymbol{w}$, we get $\boldsymbol{w} \,.\, \mathbf{x_1} = (a_1\,\mathbf{x_1} - a_2\,\mathbf{x_2}) \,.\, \mathbf{x_1} = a_1\mathbf{x_1} \,.\, \mathbf{x_1} = a_1$ because $\mathbf{x_1}$ and $\mathbf{x_2}$ are orthogonal and $\mathbf{x_1}$ is a unit vector. Hence $\boldsymbol{w} \,.\, \mathbf{x_1} = a_1 = 1$. Same reasoning yields $a_2 = 1$. So $\boldsymbol{w} = (a_1\,\mathbf{x_1} - a_2\,\mathbf{x_2}) = \mathbf{x_1} - \mathbf{x_2}$ **= [1 –1].**