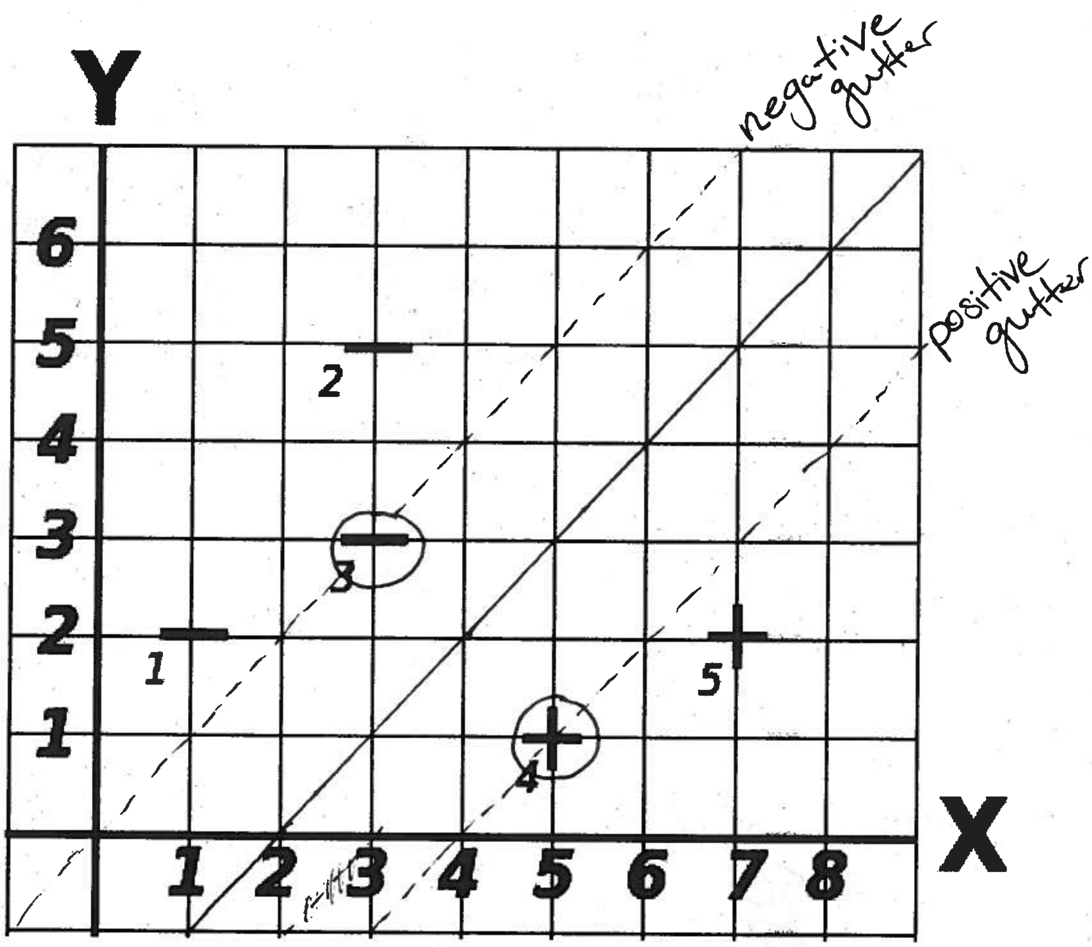
Problem 1: SVMs (50 points)

After reading too much Lord of the Rings, you wake up to find yourself in Middle Earth. You decide the most relevant thing to do is to classify the different races around you.

Part A: Distinguishing Dwarves from Humans (38 points)

You meet 2 dwarves (+) and 3 humans (-) and realize that they have distinguishing features: beard length (x) and height (y). You plot these data points on a grid. Being an expert in SVMs, you decide to start off on an epic journey of classification.



A1 (7 points)

Draw the **decision boundary** on the graph above and **clearly label positive and negative gutters, and circle all support vectors.**

What is the width of the road/margin?

A2 (12 points)

Compute $\vec{w}$ and b in the decision boundary $h(\vec{u}) = \vec{w} \cdot \vec{u} + b \geq 0$
for the SVM solution to part A1.

Show your work here.

$\vec{w}$ =

b =

A2 (10 points)

Calculate the weights (alphas) of each data point.
Show your work here.

α_1 =

α_2 =

α_3 =

α_4 =

α_5 =

A3 (9 points)

What will be the alpha of a new negative point 6 placed at (0, 6)?

What will be the alpha of a new negative point 6 placed at (0,0)?

Supposed we moved point 3 to (4, 2), how will the **magnitude of the alpha 3 change?**

Circle one:

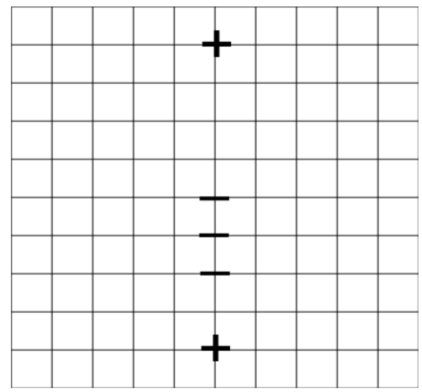
Larger Smaller Same

Part B: Distinguishing Kernels (12 points)

Back in his lab, Gandolf has been hacking on some kernels in preparation for greater classification adventures. For each of the following, indicate **YES** or **NO** whether the kernel can be used to ***perfectly*** classify the test points, and if **YES *sketch* the decision boundaries and gutters (the street) such a classifier might produce** and ***circle* which data points are support vectors**. Note that because of symmetry, more than one answer may be possible for one or more cases.

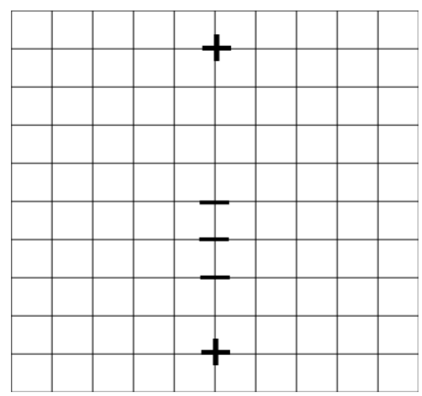
$$K(\vec{u}, \vec{v}) = \vec{u} \cdot \vec{v}$$

YES **NO**



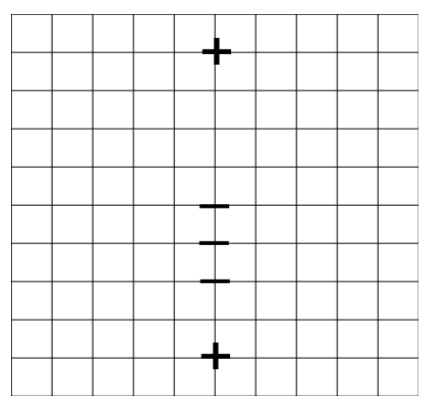
$$K(\vec{u}, \vec{v}) = (\vec{u} \cdot \vec{v} + 1)^2$$

YES **NO**



$$K(\vec{u}, \vec{v}) = e^{-(|\vec{u} - \vec{v}|)^2/2}$$

YES **NO**



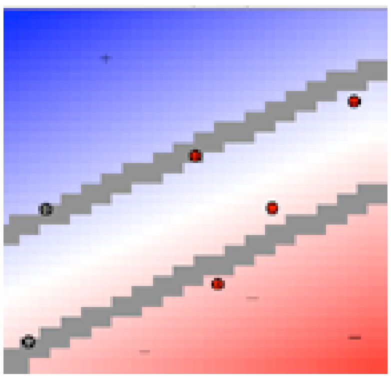
Final Exam 2002 Problem 6: Support Vector Machines (14 Points)

Part A: (2 Points)

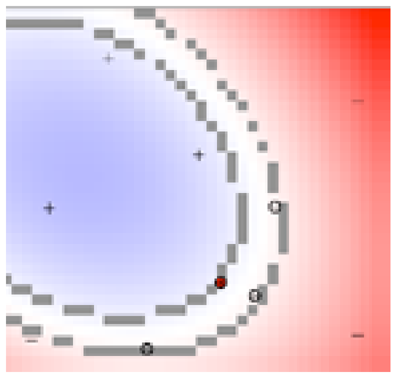
The following diagrams represent graphs of support vector machines trained to separate pluses (+) from minuses (-) for the same data set. The origin is at the lower left corner in all diagrams. Which represents the best classifier for the training data? ***See the separate color sheet for a clearer view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

Indicate your choice here:

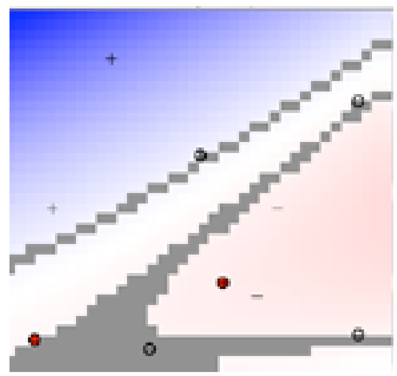
A.



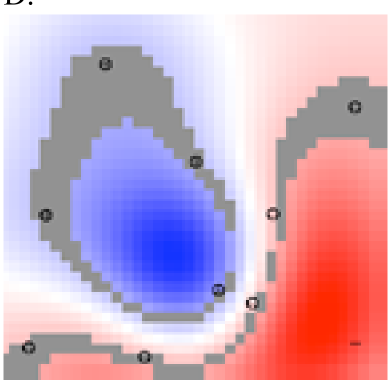
B.



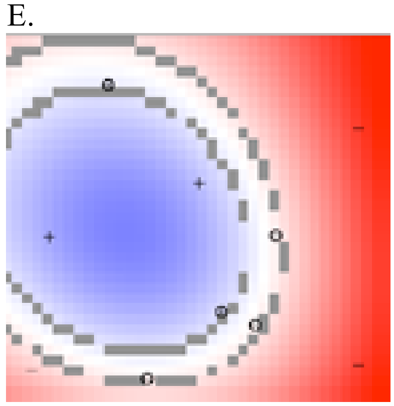
C.



D.



E.



Part B: (5 Points)

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

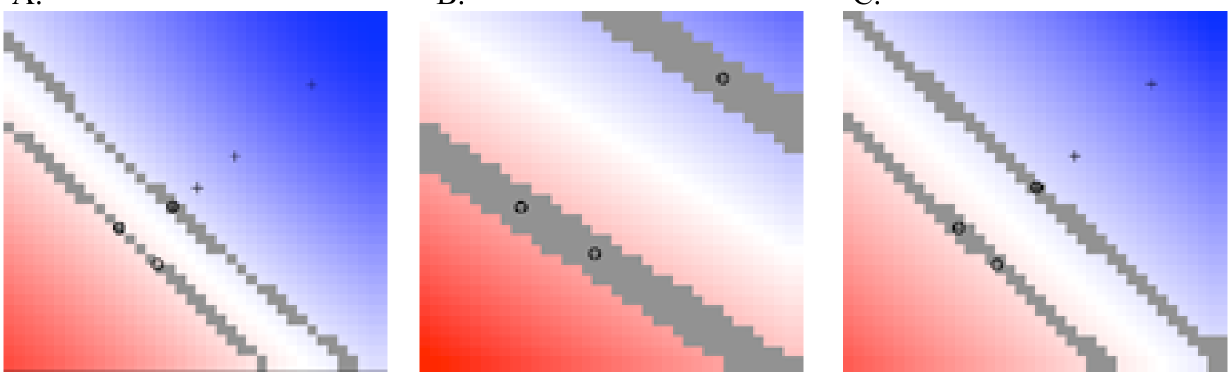
Part C: (3 Points)

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***

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 ***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 α_i or y_i. Hint: Think about the values produced by the decision rule for the support vectors, $\mathbf{x_1}$ and $\mathbf{x_2}$.

w [ ]

b [ ]

Problem 2: Boosting (50 points)

After wearing Sauron's ring for several months, Frodo is rapidly losing his sanity. He fears that the ring will interfere with his better judgement and betray him to an enemy. To ensure that he doesn't put his trust into enemy hands, he flees Middle Earth in search of a way to classify his enemies from his friends. In his travels he had heard rumors of the magic of Artificial Intelligence and has decided to hire you to build him a classifier, which will correctly differentiate between his friends and his enemies. Below is all of the information Frodo remembers about the people back in Middle Earth.

ID	Name	Friend	Species	Has Magic	Part of the Fellowship	Has/Had a ring of power	Length of hair (feet)
1	Gandalf	**Yes**	Wizard	Yes	Yes	No	2
2	Sarumon	**No**	Wizard	Yes	No	No	2.5
3	Sauron	**No**	Wizard	Yes	No	Yes	0
4	Legolas	**Yes**	Elf	Yes	Yes	No	2
5	Tree-Beard	**Yes**	Ent	No	No	No	0
6	Sam	**Yes**	Hobbit	No	Yes	No	0.25
7	Elrond	**Yes**	Elf	Yes	No	Yes	2
8	Gollum	**No**	Hobbit	No	No	Yes	1
9	Aragorn	**Yes**	Man	No	Yes	No	0.75
10	Witch-king of Angmar	**No**	Man	Yes	No	Yes	2.5

Part A: Picking Classifiers (10 points)

A1 (6 points)

The data has a high dimensionality and so rather than trying to learn an SVM in a high dimension space you think it would be a smart approach to come up with a series of 1 dimensional stubs that can be used to construct a boosting classifier. Fill in the classifier table below. Each of the different classifiers are given a unique ID and a test returns +1 (friend) if true and -1 (enemy) if false.

Classifier	Test	Misclassified
A	Species is a Wizard	2, 3, 4, 5, 6, 7, 9
B	Species is an Elf	1, 5, 6, 9
C	Species is **not** a Man	2, 3, 8, 9
D	Does **not** have magic	1, 4, 7, 8
E	Is **not** part of the Fellowship	1, 2, 3, 4, 6, 8, 9, 10
F	Has **never** owned a ring of power	2, 7
G	Hair <= 1ft	1, 3, 4, 7, 8
H	Hair <= 2 ft	3, 8
I	Friend	2, 3, 8, 10
J	Enemy	1, 4, 5, 6, 7, 9

A2 (4 points)

Looking at the results of your current classifiers, you quickly see two more good weak classifiers (make fewer than 4 errors). What are they?

Classifier	Test	Misclassified
K		1, 8, 10
L		

Part B: Build a Strong Classifier (30 points)

B1 (25 points)

You realize that many of your tests are redundant and decide to move forward using only these four classifiers: {**B**, **D**, **F**, **I**}. Run the Boosting algorithm on the dataset with these four classifiers. Fill in the weights, classifiers, errors and alphas for three rounds of boosting. In case of ties, favor classifiers that come first alphabetically.

Note: initial weights are set to be EQUAL and so 1/10 (they must add up to 1)

	Round 1		**Round 2**		**Round 3**	
w1	1/10	h_1 = F (why?)	F correct:	h_2 =		h_3 =
w2	1/10	Err = 2/10	F incorrect:	Err =		Err =
w3	1/10	α =		α =		α =
w4	1/10					
w5	1/10					
w6	1/10					
w7	1/10					
w8	1/10					
w9	1/10					
w10	1/10					
Err(B)	/10					
Err(D)	/10					
Err(F)	2/10 WHY?					
Err(I)	/10					

So we pick F as our first 'stump' - why?

B2 (5 points)

What is the resulting classifier that you obtain after three rounds of Boosting?

H(x) = Sign[(1/2 ln) * F(x) + (1/2ln) * + (1/2ln)*

Part C: Boost by Inspection (10 points)

As you become frustrated that you must have picked the wrong subset of classifiers to work with, one of the 6.034 TA's, Martin, happens to walk by and sees your answer to part A1. He reminds you why the boosting algorithm works and then tells you that there is no reason to actually run boosting on this dataset. A boosted classifier of the form:

$$H(x) = Sign[h_1(x) + h_2(x) + h_3(x)]$$

can be found which solves the problem. What three classifiers $\{h_1, h_2, h_3\}$ is Martin referring to, and why is the resulting H(x) guaranteed to classify all of the points correctly?