6.034 Quiz 4
December 7, 2011

Name	Solutions
Email	

Circle your **TA** *and* **recitation (for 1 extra credit point)**, so that we can more easily enter your score in our records and return your quiz to you promptly.

TAs
Avril Kenney
Adam Mustafa
Caryn Krakauer
Erek Speed
Gary Planthaber
Mick Taylor
Peter Brin
Tanya Kortz

Recitations
Thu. 1-2, Bob Berwick
Thu. 2-3, Bob Berwick
Thu. 3-4, Bob Berwick
Fri. 1-2, Randall Davis
Fri. 2-3, Randall Davis
Fri. 3-4, Randall Davis

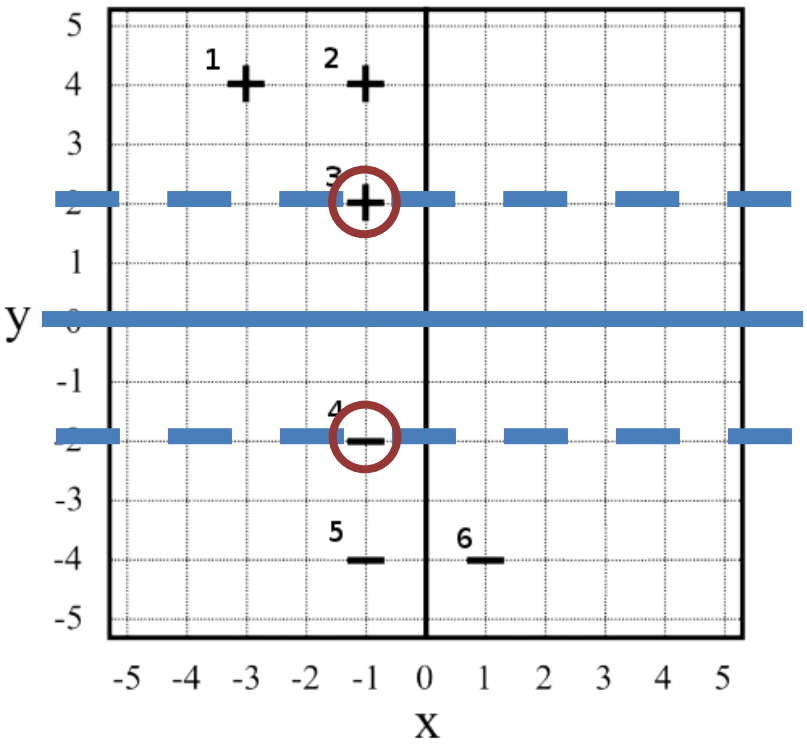
Problem	Maximum	Score	Grader
1	45		
2	45		
3	10		
Extra Credit	1		
Total	101		

There are a total of 13 pages in this quiz not including one or more tear off sheets that may be provided at the end with duplicate drawings and data. As always, open book, open notes, open just about everything, including a calculator, but no computers.

Problem 1: SVMs (45 points)

Part A: Linear kernels (22 points)

Your good friend, Khan Fusion, tells you about a binary classification technique known as a support vector machine, which appears to be all the rage among computer scientists these days. Unfortunately, Khan is a busy man, and so he could only spend 50 minutes giving you a whirlwind tour to the complex world of SVMs. Eager to see if you're on the right track, you create the following toy classification problem and solve it using an SVM to get a feeling for how it works:



A1 (6 points)

On the picture above, draw with a **heavy solid line** the optimal decision boundary (that is, where the classifier outputs exactly 0) found by an SVM using a linear kernel. Draw with a **heavy dashed line** the edges of the "street" produced by the SVM (that is, where the classifier outputs exactly +1 or -1.) **Circle all support vectors.**

A2(6 points)

Give the values that the SVM finds for the following variables:

$\mathbf{w} = \begin{bmatrix} \mathbf{0} \\ \mathbf{1/2} \end{bmatrix}$ $\alpha_1 = 0$ $\alpha_3 = 1/8$ $\alpha_5 = 0$

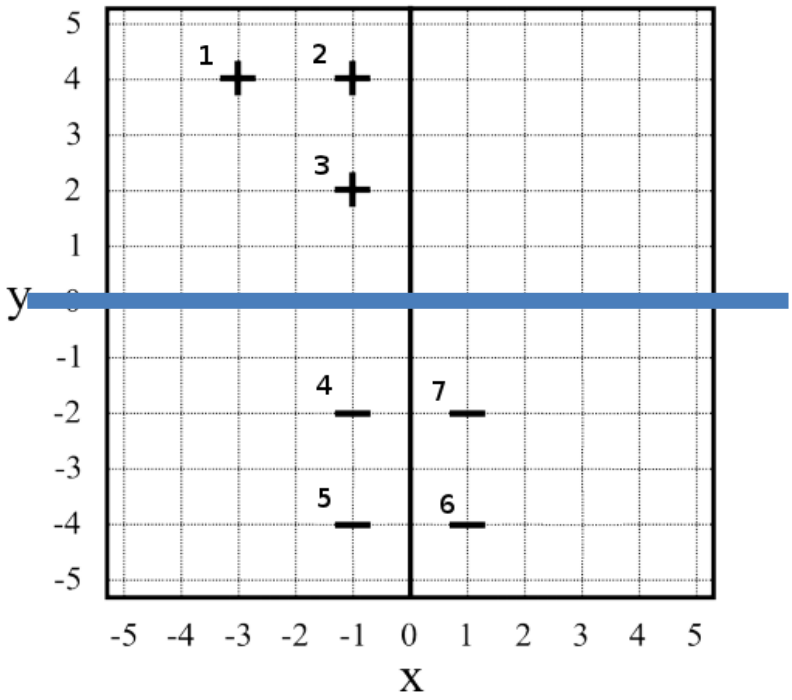
$\alpha_2 = 0$ $\alpha_4 = 1/8$ $\alpha_6 = 0$

b = 0

Show your work (for partial credit):

A3(5 points)

You now add a seventh data point to your graph as follows:



A3a Draw the decision boundary with a **heavy solid line.**

A3b What new values will the SVM find for each of the following variables?

$\mathbf{w} = \begin{bmatrix} \mathbf{0} \\ \mathbf{1/2} \end{bmatrix}$

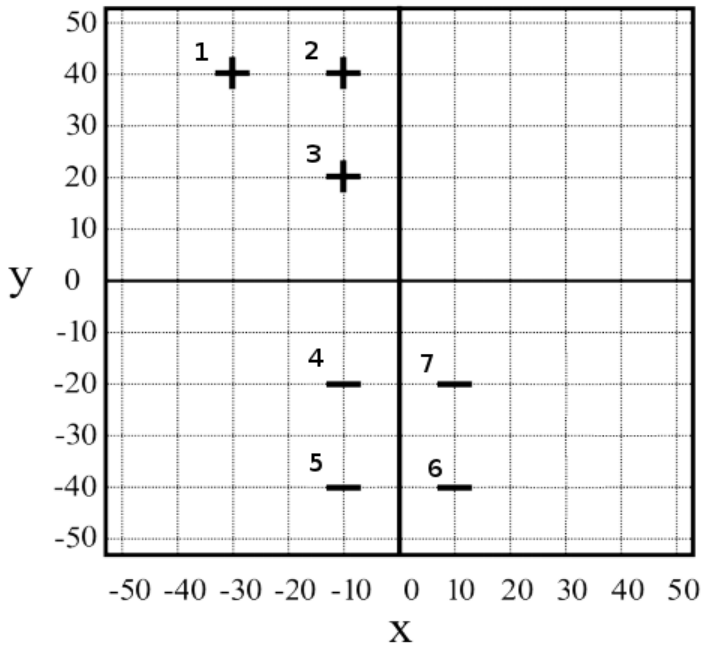
$\alpha_1 = 0$ $\alpha_3 = 1/8$ $\alpha_5 = 0$ $\alpha_7 = 0$

$\alpha_2 = 0$ $\alpha_4 = 1/8$ $\alpha_6 = 0$

b = 0

A4(5 points)

You want to see how the SVM will react to changing the scale of your problem, so you decide to multiply your data by 10 along both dimensions:



What new values will the SVM find for each of the following variables?

$\mathbf{w} = \begin{bmatrix} \mathbf{0} \\ \mathbf{1/20} \end{bmatrix}$ $\alpha_1 = 0$ $\alpha_3 = 1/800$ $\alpha_5 = 0$ $\alpha_7 = 0$

$\alpha_2 = 0$ $\alpha_4 = 1/800$ $\alpha_6 = 0$

b = 0

Part B: Higher order kernels (23 points)

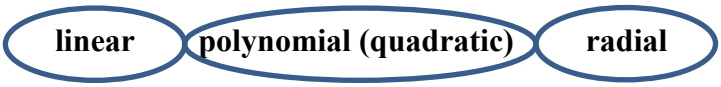
You're pretty satisfied with your understanding of SVMs at this point, so you decide to use them to tackle a slightly harder problem, certainly worthy of an MIT student's attention: given two input **integers** x and y, can we learn to predict whether their sum, x + y, will be evenly divisible by 2?

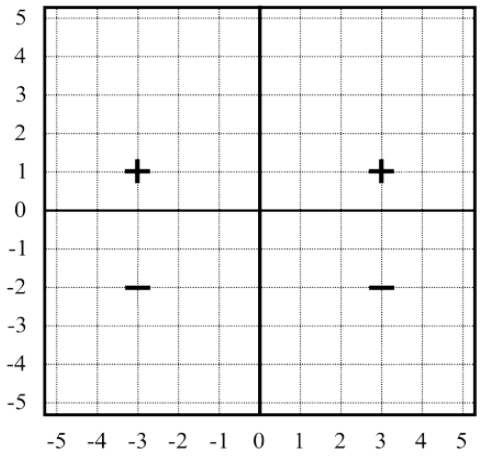
(Note: for the purpose of this problem, we consider 0 to be evenly divisible by 2.)

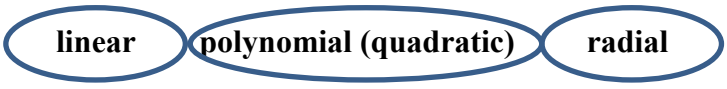
B1 (9 points)

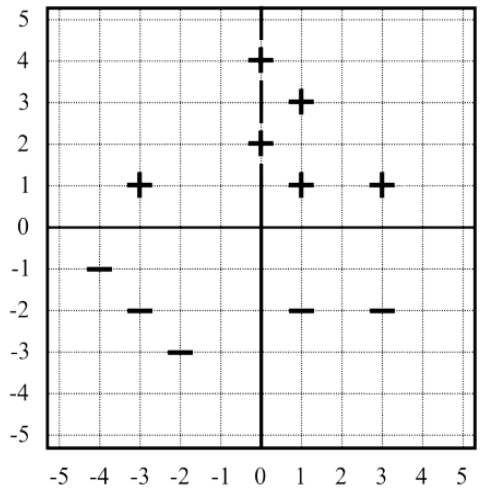
You slowly start to populate your training set by inserting data points which you manually classify.

For each of the following graphs, **circle ALL kernels that can perfectly classify the training data.** Your three options are: **linear** kernels, **quadratic polynomial** kernels, and **radial basis function** kernels.

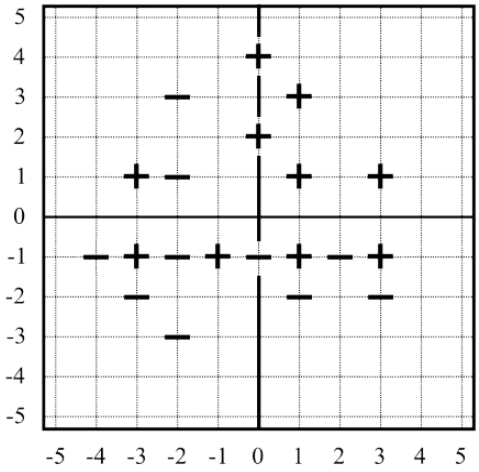








linear **polynomial (quadratic)** **radial**



B2 (4 points)

You heave a sigh of victory as you finish adding the 121st **distinct** data point to the graph above. You then run your learning algorithm with a radial basis function kernel using suitable parameters and test your classifier with two random integers between -5 and 5. Will your classifier always be able to classify this data point correctly? (**circle one**)

YES **NO**

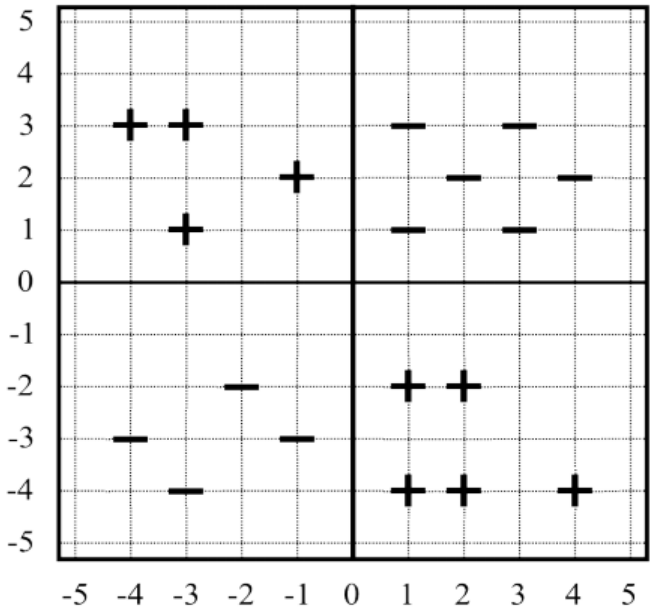
B3 (4 points)

You decide to test your classifier from B2 on random data points outside the range where x and y are between -5 and +5, **without retraining your SVM**. Will your classifier always be able to classify these data points correctly? (**circle one)**

YES **NO**

B4 (6 points)

Confused by your results above, you decide that you want to get a more intuitive sense of how kernels work, so you turn your attention to a simpler problem:



Give a transformation φ(**u**) (where **u** is a vector whose components are u_1 and u_2) that will make these data points linearly separable, and compute the kernel **K**(**u, v**) associated with the transformation.

φ(**u**): $u_1 u_2$

K(**u, v**): $\varphi(\boldsymbol{u}) \circ \varphi(\boldsymbol{v}) = u_1 u_2 v_1 v_2$

Note: A variety of solutions were accepted, but the above is one of the simplest. Another is $\boldsymbol{\varphi(u) = |u_2 - u_1|}$.

Problem 2: Adaboost (45 points)

Your friend Ben has spent the last month playing a new video game instead of attending 6.034. This weekend his girlfriend Brittney is visiting, and they want to make a character for her, but don't know what sort of character it should be. She has given him a description of her ideal character, and Ben has asked you to use your newfound knowledge of Adaboost to build a classifier and decide if she should play an elf or a non-elf.

Below is the table that Ben has compiled of training data for your classifiers. Note that +1 in the Elf column means that the person should be an elf; -1 means the person should not be an elf.

	Person	Elf	Fighting Style	Ears	Magic	Size
1	**Link**	**+1**	Ranged	Pointy	Yes	Medium
2	**Arwen**	**+1**	Melee	Pointy	No	Medium
3	**Legolas**	**+1**	Ranged	Pointy	No	Medium
4	**Dobby**	**+1**	Ranged	Pointy	Yes	Small
5	**Christmas Elf**	**+1**	None	Pointy	Yes	Small
6	**Fran**	**-1**	Ranged	Round	Yes	Medium
7	**Green Arrow**	**-1**	Ranged	Round	No	Medium
8	**Gizmo**	**-1**	None	Pointy	No	Small

Part A: Choosing Classifiers (18 points)

A1 (10 points)

Fill in the following chart of misclassified data points for the given weak classifiers.

Classifier	Test	Misclassified
a	Fighting Style = Melee	1, 3, 4, 5
b	Fighting Style = Ranged	2, 5, 6, 7
c	Fighting Style = None	1, 2, 3, 4, 8
d	Ears = Pointy	8
e	Magic = No	1, 4, 5, 7, 8
f	Size = Medium	4, 5, 6, 7
g	True	6, 7, 8

A2 (4 points)

You notice that you could add two more good, single test, weak classifiers (fewer than half of the data points are misclassified). Fill in the tests below, given the data points they misclassify.

You may test either for properties or their negations (that is, !large would be a valid test).

We have not used **h** in the table to avoid confusion with the weak classifiers that are added up to make the strong classifier, H.

Classifier	Test	Misclassified
h	Magic = Yes	2, 3, 6
i	Fighting Style ≠ None	5, 6, 7

A3 (4 points)

Ben thinks we should use all 9 of these classifiers in boosting, just to be safe. Do you agree? Circle one:

YES **NO**

If you circled NO, list the classifier(s) (by letter) that we will NOT need to use during boosting:

b, c, e, f, g

Part B: Running Adaboost (27 points)

B1 (18 points)

Now that you know which weak classifiers you'll use (either all 9 or some subset, depending on your answer to part A3), you're ready to run Adaboost. Fill in the table below for the first three rounds of Adaboost. Break ties by **REVERSE alphabetical order** of the classifier.

	Round 1	Round 2	Round 3
w1	1/8	1/14	3/66 = 1/22
w2	1/8	1/14	3/66 = 1/22
w3	1/8	1/14	3/66 = 1/22
w4	1/8	1/14	3/66 = 1/22
w5	1/8	1/14	11/66 = 1/6
w6	1/8	1/14	11/66 = 1/6
w7	1/8	1/14	11/66 = 1/6
w8	1/8	7/14 = 1/2	21/66 = 7/22
h	d	j	i
ε	1/8	3/14	17/66
α	$\frac{1}{2}\ln 7$	$\frac{1}{2}\ln\frac{11}{3}$	$\frac{1}{2}\ln\frac{49}{17}$

You can use the space below to show your work:

B2 (4 points)

What is the final classifier produced by these three rounds of Adaboost?

$$\mathrm{H(x)} = sign\left[\frac{1}{2}\ln 7 * d(x) + \frac{1}{2}\ln\frac{11}{3} * j(x) + \frac{1}{2}\ln\frac{49}{17} * i(x)\right]$$

B3 (3 points)

Does your classifier correctly classify all of the training data? Circle one:

YES **NO**

If you circled NO, list any training data points that your Adaboost classifier misclassifies:

#6: Fran

B4 (2 points)

Below is the description of Brittney's ideal character. According to your classifier, should her character be an elf?

	Elf	Fighting Style	Ears	Magic	Size
Brittney	??	Melee	Pointy	Yes	Medium

Circle one:

ELF **NOT AN ELF**

Problem 3: Representation (10 points)

Fill in the missing items in table in the most reasonable way using the transition vocabulary:

Appear, Disappear, Change (Δ), Increase (↑), Decrease (↓), not appear, not disappear, not change, not increase, and not decrease,

Assume the table is to represent aspects of the meaning found in the following. Ignore the black cells in the table. Assume time increases from left to right.

Patrick and Karen were standing together when Patrick started running. Karen chased him, caught him, stopped him, and hit him with a club.

	Interval 1	Interval 2	Interval 3	Interval 4	Interval 5
Patrick's Speed	**A or ↑**	Not Δ	Not Δ	**D**	Not A
Karen's Speed	Not A	A	Not Δ		**Not A or Not Δ ***
Distance between K and P			↓	D	Not A
Patrick's health	Not Δ	Not Δ	Not Δ	Not Δ	**↓ or Δ or D**

* We also accepted D with an explanation.