Quiz 4 Solutions

December 6, 2006

Name	
EMail	

Problem number	Maximum	Score	Grader
1	50		
2	50		
Total	100		

There are **10 pages** in this quiz, **not** including this one.

Quiz 4: SVMs and Boosting

Problem 1: Boosting Identification Trees using ADABOOST (50 points)

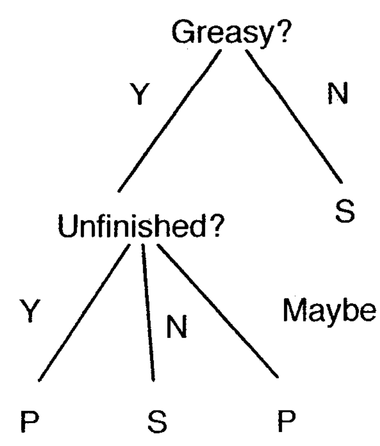
As you recall, back in 2001, you were just appointed as a research assistant at PacMan Institute of Technology, where you worked for Pinky, Winky, Binky, and Hinky on research projects.

Binky asked you to build a Ghostbot that can collect and dispose of styrofoam foodtruck containers (a very useful task around P.I.T.). They should go into two different trucks **P** (for Powerpellet Kitchen) or **S** (for Strawberry King).

You were given the following training set (repeated on a tear off sheet for your convenience at the end of this examination). Note that "Maybe" is a possible result for the Unfinished? test.

Training Ex. #	Colorful? (C)	Unfinished? (U)	Greasy? (G)	Truck (output y)
1	Y	Y	N	S (−1)
2	N	Y	N	S (−1)
3	N	Y	N	S (−1)
4	Y	Y	Y	P (+1)
5	N	N	Y	S (−1)
6	N	Maybe	Y	P (+1)

That time, you completed the project by learning the following ID tree from the training data.



Applying ADABOOST

After spending five years at P.I.T., you are significantly smarter now (or so you hope:) As a matter of fact, you have just recently learned about a great algorithm for learning strong classifiers from data called ADABOOST. You are then reminded that the outcome of that old project back in 2001 has not had the success you expected. So you decide to give ADABOOST a run at that old data.

Single-Test Classifiers (*aka* Decision Stumps)

First you define the test function as

$$\text{test}(\textit{condition}) = \begin{cases} +1, & \text{if } \textit{condition} \text{ holds,} \\ -1, & \text{otherwise.} \end{cases}$$

Mapping Problem Inputs and Outputs

Then, you map the feature variables into the vector

$$x = (\text{Colorful?}, \text{Unfinished?}, \text{Greasy?}).$$

You also map Truck output in to the output variable

$$y = \text{test}(\text{Truck} = \text{P}).$$

In other words, if Truck = P, the output is +1 and if Truck = S, the output is −1.

The Weak Learner

You then go ahead and program a simple weak learning algorithm that considers only the following simple classifiers based on the following 12 tests (6 basics plus their inverses):

Test	*condition*	Test	*condition* (inverse)
TRUE	Always true (+1)	FALSE	Always false (−1)
C=Y	Colorful? = Y	C=N	Colorful? = N
U=Y	Unfinished? = Y	U≠Y	Unfinished? ≠ Y
U=N	Unfinished? = N	U≠N	Unfinished? ≠ N
U=M	Unfinished? = Maybe	U≠M	Unfinished? ≠ Maybe
G=Y	Greasy? = Y	G=N	Greasy? = N

Note that Test TRUE (i.e., Always true) means the output of classifier TRUE is always +1, or in other words that the value of Truck is always P. Similarly, Test FALSE (i.e., Always false) means the output of classifier FALSE is always −1, or in other words that the value of Truck is always S.

If multiple tests have the same weighted error, the learner breaks ties by the test label, in order, first from TRUE to G=Y and then from FALSE to G=N, so that TRUE is the most preferred and G=N is the least preferred. In other words, the order of preference is TRUE, C=Y, U=Y, U=N, U=M, G=Y, FALSE, C=N, U≠Y, U≠N, U≠M, G=N.

Part 1.A: Only Effective Weak/Base Classifiers Needed (10 points)

Fill in the table below. (Some of the entries have been filled in for you already.)

Test	Misclassified training examples	Test	Misclassified training examples
TRUE	1, 2, 3, 5	FALSE	4, 6
C=Y	1, 6	C=N	2, 3, 4, 5
U=Y	1, 2, 3, 6	U≠Y	4, 5
U=N	4, 5, 6	U≠N	1, 2, 3
U=M	4	U≠M	1, 2, 3, 5, 6
G=Y	5	G=N	1, 2, 3, 4, 6

Which tests will **never** be used? (Cross out all that apply.)

~~All will be used~~

~~TRUE~~ C=Y ~~U=Y~~ ~~U=N~~ U=M G=Y

~~FALSE~~ ~~C=N~~ ~~U≠Y~~ U≠N ~~U≠M~~ ~~G=N~~

Part 1.B: Running and Tracking ADABOOST (20 points)

You decide to run ADABOOST for $T = 3$ rounds, using your simple weak learning algorithm. Track the execution of ADABOOST by filling in the table below the respective values you found on each round $t = 1, 2, 3$.

Training # i	$D_1(i)$	$D_2(i)$	$D_3(i)$
1	1/6	1/10	$\frac{1}{18}$
2	1/6	$\frac{1}{10}$	$\frac{1}{18}$
3	1/6	$\frac{1}{10}$	$\frac{1}{18}$
4	1/6	$\frac{1}{2}$	5/18
5	1/6	$\frac{1}{10}$	$\frac{1}{2}$
6	1/6	$\frac{1}{10}$	$\frac{1}{18}$
Weak classifier h_t	$h_1 =$ Test U=M	$h_2 =$ Test G=Y	$h_3 =$ Test C=Y
Error ϵ_t	$\epsilon_1 = \frac{1}{6}$	$\epsilon_2 = \frac{1}{10}$	$\epsilon_3 = 1/9$
Weights α'_t	$\alpha'_1 = \ln 5$	$\alpha'_2 = \ln 9$	$\alpha'_3 = \ln 8$

Part 1.C: The ADABOOST Classifier (10 points)

The final classifier output by ADABOOST is

$$H(x) = \text{sign}\Big(\underline{\ln 5}\ \text{test}\big(\text{Unfinished?} = \text{Maybe}\big) +$$
$$\underline{\ln 9}\ \text{test}\big(\underline{\text{Greasy?} = \text{Y}}\big) +$$
$$\underline{\ln 8}\ \text{test}\big(\underline{\text{Colorful?} = \text{Y}}\big)\Big).$$

Part 1.D: Comparing the ID-Tree and AdaBoost Classifiers on the Training Data (5 points)

Is your new AdaBoost classifier **worse than** your old ID tree classifier from 2001 **on the training data**? (i.e., does your new AdaBoost classifier **make more mistakes than** your old ID tree classifier from 2001 **on the training data**?)

Yes **No**

Part 1.E: Comparing the ID-Tree and AdaBoost Classifiers on New Unseen Examples (5 points)

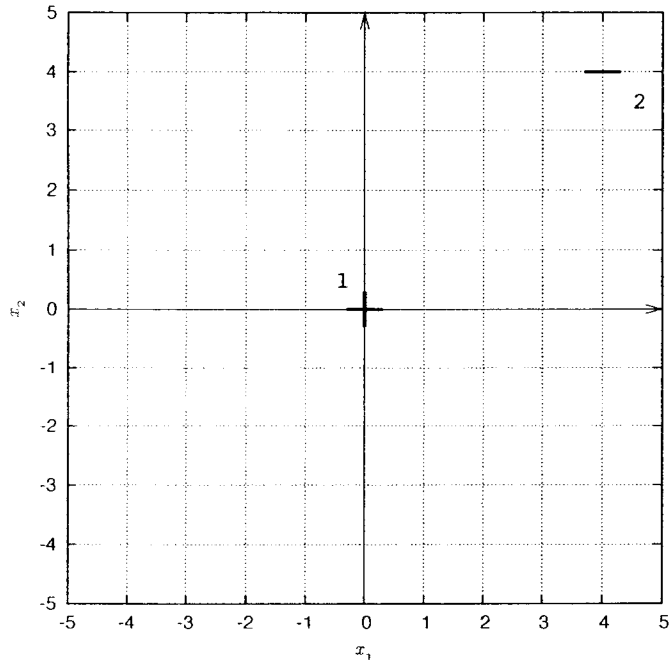
Is your new AdaBoost classifier **the same** as your old ID tree classifier from 2001? (i.e., do both classifiers agree **on every possible input** x?)

Yes **No**

Problem 2: Support Vector Machines (50 points)

Part 2.A: Working with Linear SVMs (14 points)

Consider the following data set with one positive example $\bar{x}_1 = (0, 0), y_1 = +1$ and one negative example $\bar{x}_2 = (4, 4), y_2 = -1$.



Part 2.A.1. Is the data linearly separable?

Yes   No

Part 2.A.2. If your answer to **Part 2.A.1** is **No**, then explain in a single short sentence why not?

If your answer to **Part 2.A.1** is **Yes**, then

1. provide the SVM classifier below.

$$h(\bar{x}) = \begin{cases} +1, & \text{if } \underline{-1}\, x_1 + \underline{-1}\, x_2 + \underline{4} \geq 0, \\ -1, & \text{otherwise.} \end{cases}$$

2. Also provide the weight value of the support vectors α_1 and α_2, as well as the offset-threshold value b of the SVM classifier.

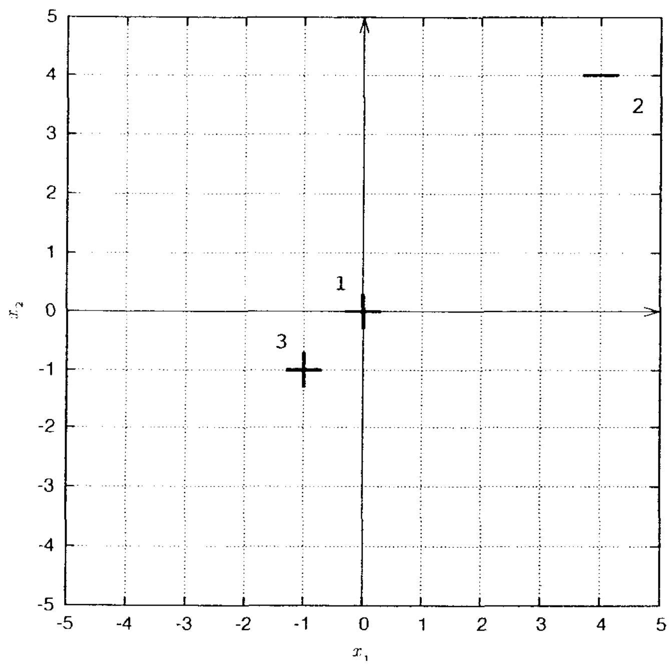
$$\alpha_1 = \underline{\frac{1}{16}}$$

$$\alpha_2 = \underline{\frac{1}{16}}$$

$$b = \underline{1}$$

Part 2.B: On Properties of Linear SVMs I (10 points)

Suppose we have an additional positive example $\bar{x}_3 = (-1, -1), y_3 = +1$.



Part 2.B.1. Which data points are support vectors? (Circle)

none 1 2 3

Part 2.B.2. If both this data set and that in **Part 2.A** are linearly separable, then

1. is the decision boundary for this data set different than for **Part 2.A**?

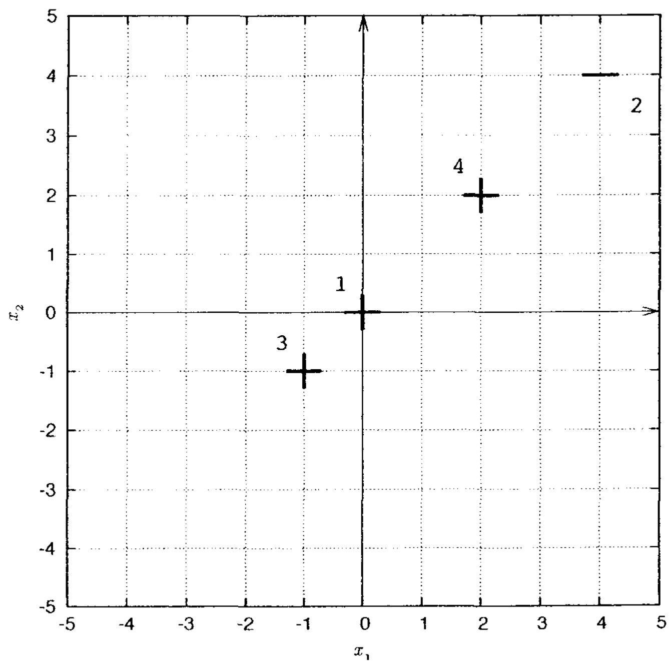
   Yes **No**

2. Relative to the support vectors for the data set in **Part 2.A**, the weights (α values) of the support vectors for the data set in this part are

   the same **smaller** **larger.**

Part 2.C: On Properties of Linear SVMs II (10 points)

Suppose we have an additional positive example $\bar{x}_4 = (2, 2), y_4 = +1$.



Part 2.C.1. Which data points are support vectors? (Circle)

none 1 2 3 4

Part 2.C.2. If both this data set and that in **Part 2.B** are linearly separable, then

1. is the decision boundary for this data set different than for **Part 2.B**?

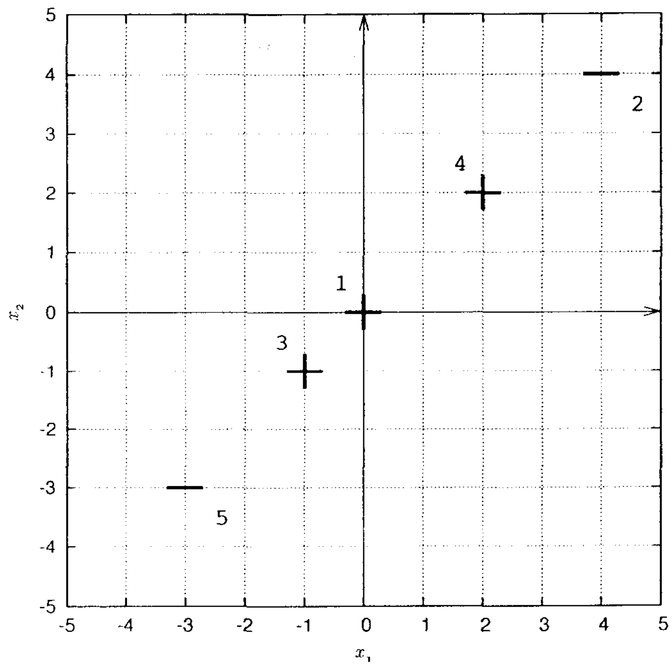
   Yes **No**

2. Relative to the support vectors for the data set in **Part 2.B**, the weights (α values) of the support vectors for the data set in this part are

   the same **smaller** **larger.**

Part 2.D: On the Power of SVMs (10 points)

Suppose we have an additional negative example $\bar{x}_5 = (-3, -3), y_5 = -1$.



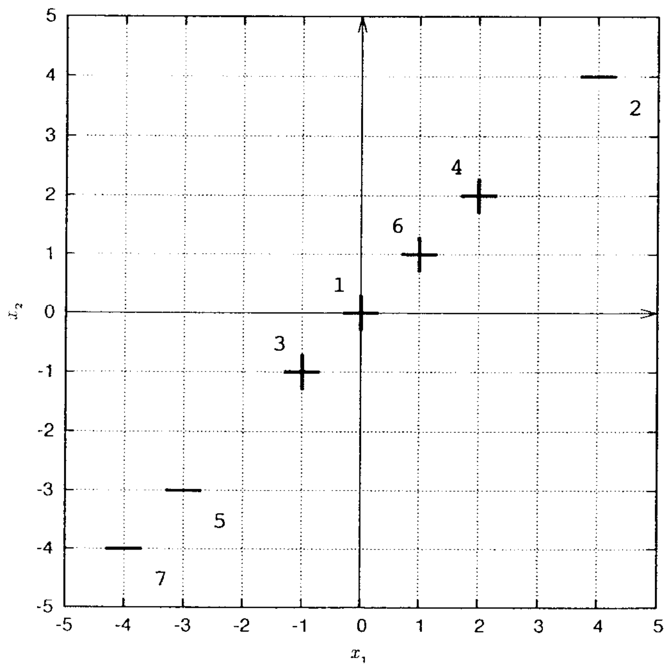
Circle the number(s) to the left side of the kernels described in the enumerated list below that can separate the data.

1. linear: $K(\bar{u}, \bar{v}) = \bar{u} \cdot \bar{v}$
2. polynomial of degree $n \geq 2$: $K(\bar{u}, \bar{v}) = (1 + \bar{u} \cdot \bar{v})^n$
3. Radial Basis Function / Gaussian with sufficiently small scale parameter σ: $K(\bar{u}, \bar{v}) = e^{-\frac{\|\bar{u}-\bar{v}\|}{2\sigma^2}}$
4. none

Part 2.E: On Data Transformations and Kernels (6 points)

NOTE: It is possible to solve this problem without solving for the α's.
Hint: Think of the definition of a kernel.

Suppose we have two additional examples: one positive example $\bar{x}_6 = (+1, +1), y_6 = +1$, and one negative example $\bar{x}_7 = (-4, -4), y_7 = -1$.



Consider using the following kernel:

$$K(\bar{u}, \bar{v}) = 2\|\bar{u}\|\|\bar{v}\|.$$

1. Find the SVM classifier:

$$h(\bar{x}) = \begin{cases} +1, & \text{if } -2\, x_1^2 + \underline{\ 0\ }\, x_1 x_2 + \underline{\ -2\ }\, x_2^2 + \underline{\ 0\ }\, x_1 + \underline{\ 0\ }\, x_2 + \underline{\ 25\ } \geq 0, \\ -1, & \text{otherwise.} \end{cases}$$

2. Which data points are support vectors? (Circle)

none 1 2 3 4 5 6 7