ADABOOST PRACTICE QUESTIONS

1. Suppose the training data is as follows:

index:	0	1	2	3	4	5	6	7	8	9
x value:	0	1	2	3	4	5	6	7	8	9
y value:	1	1	1	-1	-1	-1	1	1	1	-1

x is the attribute and y is the output (target) value.

The weak learner produces hypotheses of the form: $x < v$ or $x > v$, where v is a threshold parameter designed to minimize the probability of training error.

<u>Solution Hints:</u>

<u>Round1:</u>

Start with equal weight distribution i.e. $D_1(i) = 0.1$ for i = 0 to 9

The best weak hypothesis:

if x > 2.5 then class = -1

else if x <= 2.5 class = 1

The training error? 3 points with weight 0.1 each are misclassified, so total is 0.3

$$\epsilon_1 = \frac{0.3}{1} = 0.3$$

$$\alpha_1 = \frac{1}{2}\ln\left(\frac{1-\epsilon_1}{\epsilon_1}\right) = 0.424$$

Updated weight for correctly classified points:

$$D_2(i) = \frac{0.1\times\exp(-\alpha_1)}{Z_1} = \frac{0.065}{Z_1}$$

Updated weight for incorrectly classified points:

$$D_2(i) = \frac{0.1\times\exp(\alpha_1)}{Z_1} = \frac{0.153}{Z_1}$$

There are 7 correctly classified and 3 incorrectly classified, so total weights:

$$= 7\times\frac{0.065}{Z_1} + 3\times\frac{0.153}{Z_1}$$
$$= \frac{0.914}{Z_1}$$

Total weight should be 1, so $Z_1 = 0.914$.

Correctly classified weight = $\frac{0.065}{0.914} = 0.071$

Incorrectly classified weight = $\frac{0.153}{0.914} = 0.167$

The weight distribution after round 1:

Index	0	1	2	3	4	5	6	7	8	9
x	0	1	2	3	4	5	6	7	8	9
y	1	1	1	-1	-1	-1	1	1	1	-1
weight	0.071	0.071	0.071	0.071	0.071	0.071	0.167	0.167	0.167	0.071

What's the best function after round1?

$h_1 = \alpha_1 I(x < 2.5)$ where $I(x)$ returns 1 if x is true and -1 if x is false.

$h_1 = 0.434\ I(x < 2.5)$

Round2:

Now, you need to find best threshold with new set of weights.
If threshold is 2.5, error = 3*0.167 = 0.50
If threshold is 5.5, error = 3 on left + 1 on right = 3 * 0.071 + 1 * 0.071 = 0.284
If threshold is 8.5, error = 3 on left + 1 on right = 3 * 0.071 = 0.213

The 2nd round hypothesis is:

$h_2 = \alpha_2 I(x < 8.5)$ where $I(x)$ returns 1 if x is true and -1 if x is false.

$$\epsilon_2 = \frac{0.213}{1} = 0.213$$

$$\alpha_2 = \frac{1}{2}\ln\left(\frac{1-\epsilon_2}{\epsilon_2}\right) = 0.653$$

Updated weight for correctly classified points:

for i = 0, 1, 2, 9

$$D_3(i) = \frac{0.071 \times \exp(-\alpha_2)}{Z_2} = \frac{0.037}{Z_2}$$

for i = 6, 7, 8

$$D_3(i) = \frac{0.167 \times \exp(-\alpha_2)}{Z_2} = \frac{0.087}{Z_2}$$

Updated weight for incorrectly classified points:

for i = 3, 4, 5

$$D_2(i) = \frac{0.071 \times \exp(\alpha_2)}{Z_2} = \frac{0.136}{Z_2}$$

Since the sum of weights should be 1, Z_2 = 0.817

Index	0	1	2	3	4	5	6	7	8	9
x	0	1	2	3	4	5	6	7	8	9
y	1	1	1	-1	-1	-1	1	1	1	-1
weight	0.045	0.045	0.045	0.166	0.166	0.166	0.106	0.106	0.106	0.045

What's the best function after round2?

$h_2 = \alpha_2 I(x < 8.5)$ where $I(x)$ returns 1 if x is true and -1 if x is false.

$h_2 = 0.653 \ I(x < 8.5)$

If I combine h1 and h2, do I get perfect classification?

H(x) = sign(0.434 $I(x < 2.5)$ + 0.653 $I(x < 8.5)$)

Let's see what happens to points 3 to 8 i.e. i = 3 to 8

H_i(x) = sign(0.434 (-1) + 0.653 (1)) = 1 => Error for points 3, 4, and 5 so need round 3.

Round3:

Now, you need to find best threshold with new set of weights.
If threshold is 2.5, error = 3*0.106 = 0.318
If threshold is 5.5, error = 3 on left + 1 on right = 3 * 0.045 + 1 * 0.045 = 0.180
If threshold is 8.5, error = 6 on left = 3 * 0.045 + 3 * 0.106 = 0.453

The 3rd round hypothesis is:

$h_3 = \alpha_3 I(x > 5.5)$ where $I(x)$ returns 1 if x is true and -1 if x is false.

$$\epsilon_3 = \frac{0.180}{1} = 0.180$$

$$\alpha_3 = \frac{1}{2}\ln(\frac{1-\epsilon_3}{\epsilon_3}) = 0.758$$

Updated weight for correctly classified points:

for i = 3, 4, 5

$$D_4(i) = \frac{0.166\times\exp(-\alpha_3)}{Z_3} = \frac{0.078}{Z_3}$$

for i = 6, 7, 8

$$D_4(i) = \frac{0.106\times\exp(-\alpha_3)}{Z_3} = \frac{0.050}{Z_3}$$

Updated weight for incorrectly classified points:

for i = 0, 1, 2, 9

$$D_3(i) = \frac{0.045\times\exp(\alpha_3)}{Z_3} = \frac{0.096}{Z_3}$$

Since the sum of weights should be 1, Z3 = 0.768

Index	0	1	2	3	4	5	6	7	8	9
x	0	1	2	3	4	5	6	7	8	9
y	1	1	1	-1	-1	-1	1	1	1	-1
weight	0.125	0.125	0.125	0.101	0.101	0.101	0.065	0.065	0.065	0.125

What's the best function after round3?

$h_3 = \alpha_3 I(x > 5.5)$ where $I(x)$ returns 1 if x is true and -1 if x is false.

$h_3 = 0.758 \ I(x > 5.5)$

If I combine h1, h2, and h3, do I get perfect classification? Let's check:

H(x) = sign($0.434 \ I(x < 2.5) + 0.653 \ I(x < 8.5) + 0.758 \ I(x > 5.5)$)

Index	0	1	2	3	4	5	6	7	8	9
x	0	1	2	3	4	5	6	7	8	9
y	1	1	1	-1	-1	-1	1	1	1	-1
weight	0.125	0.125	0.125	0.101	0.101	0.101	0.065	0.065	0.065	0.125
H(x)	1	1	1	-1	-1	-1	1	1	1	-1

Perfect classification!! Final hypothesis is:

H(x) = sign($0.434 \ I(x < 2.5) + 0.653 \ I(x < 8.5) + 0.758 \ I(x > 5.5)$)

2. Admissions data for MIT is presented below:

Table 2: MIT Admissions Training Data

ID	Name	**Admit/Deny**	# of High School Detentions	SAT
1	Andrew	Deny	3	2050
2	Burt	Admit	1	2200
3	Charlie	Admit	2	2090
4	Derek	Deny	4	2230
5	Erica	Admit	5	2330
6	Faye	Deny	6	2220
7	Greg	Admit	6	2390
8	Helga	Admit	7	2320
9	Ivana	Deny	8	2330
10	Jan	Deny	8	2090

The two attributes are detentions (x1), and SAT scores (x2).
The predicted output is admit/deny (y).

You decide to make weak hypothesis of the form h1: x1 > v or h2: x2 > v

Run Adaboost algorithm on this data to get perfect training classification.

What is the final hypothesis?

Note: If you would like to visualize this data, here is the R code:

```
detentions<-c(3, 1, 2, 4, 5, 6, 6, 7, 8, 8)
sat<-c(2050, 2200, 2090, 2230, 2330, 2220, 2390, 2320, 2330, 2090)
admit<-c(0, 1, 1, 0, 1, 0, 1, 1, 0, 0)
mitAdmissions<-data.frame(detentions,sat,admit)
library(ggplot2)
qplot(detentions, sat, color=admit, data=mitAdmissions)
```