Question 1:

No	X	Y	Initial Weights	$\hat{y}$	Correct ?	Updated weights
1	1.0	1	0.072	1	yes	0.0537
2	2.0	1	0.072	1	yes	0.0537
3	3.0	1	0.072	1	yes	0.0537
4	4.0	-1	0.072	-1	Yes	0.0537
5	5.0	-1	0.072	-1	Yes	0.0537
6	6.0	-1	0.072	-1	yes	0.0537
7	7.0	1	0.167	1	yes	0.1249
8	8.0	1	0.167	-1	No	0.2492
9	9.0	1	0.167	-1	No	0.2492
10	10.0	-1	0.072	-1	yes	0.0537

If y = $\hat{y}$ ‘yes’ else ‘no’

Step 1:

Finding the error rate

Error rate -> € =w.($\hat{y} \neq y$)

There fore, error rate

(E) = (0.072, 0.072, 0.072, 0.072, 0.072, 0.072,0.167,0.167,01.67,0.72)

(0,0,0,0,0,0,0,0,1,1,0)

Error rate (E) = (0.072*0)+ (0.072*0)+ (0.072*0)+ (0.072*0)+ (0.072*0)+ (0.072*0)+(0.167*0)+(0.167*1)+(0.0167*1)+(0.072*0)

= 0+0+0+0+0+0+0+0.167+0.167+0 = 0.334

Step 2:

Find the coefficient

(αj) = 0.5 log $\frac{1-e}{e}$

=0.5 log($\frac{1-0.334}{0.334}$)

= 0.3450

Step 3:

Now, we update weights

(w) = w * exp(-αj * $\hat{y}$ * y)

W1 = 0.072 * exp(-0.3450 * 1*1) = 0.0509

W2 = 0.072 * exp(-0.3450 * 1*1) = 0.0509

W3 = 0.072 * exp(-0.3450 * 1*1) = 0.0509

W4 = 0.072 * exp(-0.3450 * -1*-1) = 0.0509

W5 = 0.072 * exp(-0.3450 * -1*-1) = 0.0509

W6 = 0.072*exp(-0.3450 * -1*-1) =0.0509

W7 = 0.167*exp(-0.3450*1*1)= 0.1182

W8 = 0.167*exp(-0.3450*-1*1)= 0.2358

W9 = 0.167*exp(-0.3450*-1*1)= 0.2358

W10 = 0.072*exp(-0.3450 * -1*-1) =0.0509

Step 4:

Now we need to normalize weights, for that ∑iWi = 7*(0.0509)+0.1182+2*(0.23580)

=0.9461

Now we normalize

W = 0.0509/0.9461 =0.0537, 0.1182/0.9461 =0.1249, 0.2358/0.9461=0.2492

So that we can update the table with updated weights.

Performance analysis:

classifier: Manmographic_Masses data

No	Dtree	SVM	Bagging classifier	Random Forest	ADA Boost
Accuracy(training)	0.9457831325301205	0.7996987951807228	0.9457831325301205	0.9457831325301205	0.9457831325301205
Accuracy(testing)	0.7891566265060241	0.8072289156626506	0.8253012048192772	0.8313253012048193	0.8012048192771084

When I decreased the no.of estimators to 300 there is no much difference in the performance expect bagging classifier accuracy was slightly reduced when I changes base estimator to SVC

Accuracy sore for testing data decision tree 0.7891566265060241

Accuracy sore for training decision tree 0.9457831325301205

Accuracy sore for testing data SVM 0.8072289156626506

Accuracy sore for training SVM 0.7996987951807228

Accuracy sore for testing data bagging classifier 0.8373493975903614

Accuracy sore for training data bagging classifier 0.8328313253012049

Accuracy sore for testing data random 0.8433734939759037

Accuracy sore for training data random tree 0.9457831325301205

Accuracy sore for testing data ADA 0.7891566265060241

Accuracy sore for training ADA 0.9457831325301205

Digits data

No	Dtree	SVM	Bagging classifier	Random Forest	ADA Boost
Accuracy(training)	1.0	1.0	1.0	1.0	1.0
Accuracy(testing)	0.8583333333333333	0.9722222222222222	0.9583333333333334	0.9777777777777777	0.875

When I changed the no.of estimators to 300 and put base estimator as SVC() in the parameters. There is no much change in the performance.

Accuracy sore for testing data decision tree 0.8583333333333333

Accuracy sore for training decision tree 1.0

Accuracy sore for testing data SVM 0.9722222222222222

Accuracy sore for training SVM 1.0

Accuracy sore for testing data bagging classifier 0.9805555555555555

Accuracy sore for training data bagging classifier 0.9965205288796103

Accuracy sore for testing data random 0.9777777777777777

Accuracy sore for training data random tree 1.0

Accuracy sore for testing data ADA 0.875

Accuracy sore for training ADA 1.0

Finally, I have observed that if no.of estimators is changing then there are slight changes in the performance, Having highest no.of estimators can give the better performance.