Week 11

Question 1

Statement

In each round of AdaBoost, the weight for a particular training observation is increased going from round t to round $t+1$ if the observation was...

Options

(a)

classified incorrectly by the weak learner trained in round t

(b)

classified correctly by the weak learner trained in round t

(c)

classified incorrectly by a majority of the weak learners trained up to round t

(d)

classified correctly by a majority of the weak learners trained up to round t

Answer

(a)

solution

Since in AdaBoost, we increase the weight of the incorrectly classified points for the next bag, the option (a) is correct.

Question 2

Statement

Which of the following statements are true about bagging?

Options

(a)

In general, the final model has a higher bias than the individual learners.

(b)

In general, the final model has less bias than the individual learners.

(c)

In general, the final model has a higher variance than the individual learners.

(d)

In general, the final model has less variance than the individual learners.

Answer

(d)

Solution

Bagging on high variance models will reduce the variance without increasing the bias.
There is always a tradeoff between bias and variance. And reducing variance may cost increment in the bias. But bagging on high variance and low bias models reduces the variance without making the predictions biased.

Question 3

Statement

Is the following statement true or false?

If a point lies between the supporting hyperplanes in the soft-margin SVM problem, it always pays a positive bribe and plays a role in defining w^*.

Options

(a)

True

(b)

False

Answer

(a)

Solution

If a point lies between the supporting hyperplanes, it satisfies the following:

$$w^{*T} x_i y_i < 1 \tag{1}$$

Using the 1^{st} constraint,

$$
\begin{aligned}
1 - {w^*}^T x_i y_i - \xi_i^* &\leq 0 \\
\Rightarrow \xi_i^* &\geq 1 - {w^*}^T x_i y_i \\
\text{from } 1\text{, we can conclude that} & \\
\xi_i^* &> 0
\end{aligned}
$$

It implies that if a point lies between the supporting hyperplanes in the soft-margin SVM problem, it always pays a positive bribe

Using the CS 2,

$$
\begin{aligned}
\beta_i^* \xi_i^* &= 0 \\
\Rightarrow \beta_i^* &= 0 \quad (\text{as } \xi_i^* > 0)
\end{aligned}
$$

It implies that $\alpha_i^* = C$ and therefore If a point lies between the supporting hyperplanes in the soft-margin SVM problem, it plays a role in defining w^*.

Question 4

Statement

Is the following statement true or false?

If i^{th} point in soft-margin SVM pays a non-zero bribe $(\xi_i > 0)$, then the value of α_i is C.

Options

(a)

True

(b)

False

Answer

(a)

Solution

Using the CS 2,

$$
\begin{aligned}
\beta_i^* \xi_i^* &= 0 \\
\Rightarrow \beta_i^* &= 0 \quad (\text{as } \xi_i^* > 0)
\end{aligned}
$$

It implies that $\alpha_i^* = C$

Common data for question 5 and 6

Statement

Consider that an AdaBoost model is trained on the following binary classification dataset.

x_1	x_2	**Label** (y)
3.7	2	0
2.0	2	0
5	4	1
2.9	5	0
4.1	6	1

The first decision stump was created using the question $x_2 < 4$ or not. The error of a decision stump is defined as the proportion of misclassified points.

Question 5

Statement

Find the value of α_0. Notation is defined as per lecture.

Options

(a)

$\ln 2$

(b)

$\ln 4$

(c)

$\ln(4/5)$

(c)

$\ln(\sqrt{3/2})$

Answer

(a)

Solution

If we split the root node as per question $x_2 < 4$ or not, the left node will contain the points $(3.7, 2), (2.0, 0)$ and the labels of these points are $0, 0$ respectively. Therefore, the prediction in left node will be 0 (the majority class).

Similarly, in the right nodes, labels will be $1, 0$, and 1 and the prediction will be 1 (the majority class).

Only one point $(2.9, 5)$ is misclassified.

Therefore, error is $e = \frac{1}{5}$

$$\begin{aligned}\alpha_0 &= \ln \sqrt{\frac{1-e}{e}} \\ &= \ln \sqrt{\frac{1-1/5}{1/5}} \\ &= \ln 2\end{aligned}$$

Question 6

Statement

How will the weight corresponding to the last example change for creating the next stump?

Options

(a)

It will increase

(b)

It will decrease

Answer

(b)

Since the last example is correctly classified, its weight will decrease.

Question 7

Statement

A strong learner L is formed as per the AdaBoost algorithm by three weak learners L_1, L_2, and L_3. Their performance/weights (α) are $1, 0.4$, and 1.6, respectively. For a particular point, L_1 and L_2 predict that its label is positive, and L_3 predicts that it's negative. What is the final prediction the learner L makes on this point? Enter 1 or -1.

Answer

-1 No range is required

Solution

For the final prediction, we have

$$\alpha_1 h_1(x) + \alpha_2 h_2(x) + \alpha_3 h_3(x) = 1 + 0.4 + 1.6(-1) = -0.2 < 0$$

Therefore, prediction will be -1

Question 8

Statement

Which of the following options is correct? Select all that apply.

Options

(a)

In bagging, typically around $\frac{1}{3}rd$ data points remain unselected in bags if the number of data points is large.

(b)

Each weak leaner has equal importance in making the final prediction in Bagging.

(c)

Each weak leaner has equal importance in making the final prediction in AdaBoost.

(d)

Generally, weak learners in the random forest tend to overfit.

Answer

(a), (b), (d)

Solution

The probability that a point will not be selected in any one pick will be $(1 - 1/n)$.

The probability that a point will not be selected in n picks will be $(1 - 1/n)^n$.

as $n \to \infty, \ \ (1 - 1/n)^n \to 0.33$

that is In bagging, typically around $\frac{1}{3}rd$ data points remain unselected in bags if the number of data points is large.

In bagging, each learner has equal importance in making the final prediction as the majority of all the predictions are taken into account, and therefore each prediction counts.

But in AdaBoost, the weighted average is taken into account, and the estimator which has a higher value of α will have a higher importance in making the final prediction.

In the random forest, overfit models are preferred as they have high variance and low bias.

Common data for questions 9, 10, and 11

Statement

We have trained four models in the same dataset with different hyperparameters. In the following table, we have recorded the training and testing errors for each of the models.

Model	Training error	Test error
1	0.2	1.8

Model	Training error	Test error
2	1.0	1.1
3	0.5	0.7
4	1.9	2.3

Question 9

Statement

Which model tends to underfit?

Options

(a)

Model 1

(b)

Model 2

(c)

Model 3

(d)

Model 4

Answer

(d)

Solution

Model 4 has high training error as well as high test error. It means that model 4 as high variance and high bias and tends to underfit.

Question 10

Statement

Which model tends to overfit?

Options

(a)

Model 1

(b)

Model 2

(c)

Model 3

(d)

Model 4

Answer

(a)

Model 1 has less training error as well as high test error. It means that model 4 as high variance and low bias and tends to overfit.

Question 11

Statement

Which model would you choose?

Options

(a)

Model 1

(b)

Model 2

(c)

Model 3

(d)

Model 4

Answer

(c)

Model 3 has less training and test error and therefore, it is most preferred.