Week 11

Question 1

Statement

In a random forest model, let $p < d$ be the number of randomly selected features that are used to identify the best split at any node of a tree. Which of the following is true? (d is the total number of features)

Options

(a)

Increasing p reduces the correlation between any two trees in the forest.

(b)

Decreasing p reduces the correlation between any two trees in the forest.

(c)

Increasing p increases the performance of individual trees in the forest.

(d)

Decreasing p increases the performance of individual trees in the forest.

Answer

(b), (c)

Solution

If we increase the number of randomly selected features in a random forest model, classifiers tend to have a similar structure and the correlation between them will be higher. But decreasing the number of features may lead to a wide variety of trees and the correlation between them will decrease.

Increasing the number of features increases the performance of individual trees as it considers the most number of features to split the node.

Question 2

Statement

A dataset was generated as per the following process:

$$y = a + bx + cx^2 + dx^3 + ex^4 + \varepsilon$$

where $a, b, c, d,$ and e are the constants and the error ε follows Gaussian distribution with mean 0 and a fixed variance. The following two models were fit on the same dataset:

- model 1: $y = w_0 + w_1 x_1$
- model 2: $y = w_0 + w_1 x + w_2 x^2 + \ldots + w_8 x^8 + w_9 x^9$

 Select the correct options.

Options

(a)

Model 1 has a higher bias than model 2.

(b)

Model 1 has a higher variance than model 2.

(c)

Model 1 is more likely to overfit.

(d)

Model 2 is more likely to overfit.

(e)

Model 1 is more likely to underfit.

(f)

Model 2 is more likely to underfit.

Answer

(a), (d), (e)

Solution

Model 1 has a less complex structure to capture the real structure of the dataset and therefore, it has a higher bias than model 2 as model 2 will try to fit each and every point in the dataset.

If we make a slight change in the dataset, model 2 may deviate a lot as it always tries to capture noise in the dataset, and therefore model 2 has a higher variance than model 1.

Model 1 has high bias and low variance and it tends to underfit.

Model 2 has low bias and high variance and therefore it tends to overfit.

Question 3

Statement

How does bagging help in improving the classification performance?

Options

(a)

It helps in reducing bias

(b)

It helps in reducing variance

(c)

Bagging is inefficient if the classifiers are fully uncorrelated (independent).

(d)

Bagging is inefficient if the classifiers are fully correlated.

Answer

(b), (d)

Solution

In bagging, classifiers having high variance are trained, and their aggregate outputs are considered as the final model. The aggregation results in reducing the variances of each classifier.

If the classifiers are fully correlated, it means on average their output is the same and, therefore aggregating them does not make much difference.

Question 4

Statement

Is the following statement true or false?

If $\alpha_i^* = 0$ in the soft-margin SVM problem, then the ith data point is correctly classified by the optimal w^*.

Options

(a)

True

(b)

False

Answer

(a)

Solution

Given that $\alpha_i^* = 0$.

Since $\alpha_i^* + \beta_i^* = C, \Rightarrow \beta_i^* = C$

Using 2nd CS condition, we have

$$\beta_i^* \xi_i^* = 0$$
$$\Rightarrow \xi_i^* = 0 \qquad (\because \beta_i^* = C) \tag{1}$$

Now the first constraint, we have

$$1 - w^{*T} x_i y_i - \xi_i^* \leq 0$$
$$\Rightarrow 1 - w^{*T} x_i y_i \leq 0$$
$$\Rightarrow w^{*T} x_i y_i \geq 1$$

It implies that i^{th} data point is correctly classified by w^*.

Question 5

Statement

Assume that the first decision stump of the AdaBoost algorithm misclassifies 30 data points out of 100 data points. What will be the weights assigned to the incorrectly classified points for sampling the data points to train the second decision stump? Assume that error is defined as the proportion of incorrectly classified examples by the decision stump.

Options

(a)

$$\frac{\sqrt{7/3}}{100}$$

(b)

$$\frac{\sqrt{3/7}}{100}$$

(c)

$$\frac{\sqrt{7/3}}{30\sqrt{7/3} + 70\sqrt{3/7}}$$

(d)

$$\frac{\sqrt{3/7}}{30\sqrt{3/7} + 70\sqrt{7/3}}$$

Answer

(c)

Solution

Weights assigned for creating the 1st bag are given by

$$D_0(i) = \frac{1}{n} = \frac{1}{100}$$

the error by the first decision stump is

$$error = 0.3$$

Therefore

$$\alpha = \ln\left(\sqrt{\frac{1 - 0.3}{0.3}}\right) = \ln\left(\sqrt{\frac{7}{3}}\right)$$

The weight for the incorrectly classified points will be increased by

$$w_1(i) = w_0(i)e^{\alpha} \quad \text{the i for which } h_i(x) \neq y_i$$
$$= \frac{\sqrt{7/3}}{100}$$

The weight for the correctly classified points will be decreased by

$$w_1(i) = w_0(i)e^{-\alpha} \quad \text{the i for which } h_i(x) = y_i$$
$$= \frac{\sqrt{3/7}}{100}$$

There are 30 incorrectly classified points and 70 correctly classified points. We will normalize these weights to sum over all points equal to 1. For that, we will divide each weight with the sum of all weights.

That is

For incorrectly classified points weights will be

$$\frac{\frac{\sqrt{7/3}}{100}}{\frac{30\sqrt{7/3}}{100} + \frac{70\sqrt{3/7}}{100}}$$
$$= \frac{\sqrt{7/3}}{30\sqrt{7/3} + 70\sqrt{3/7}}$$

Common data for questions 6 and 7

Statement

You have been given a dataset in 1-d space, which consists of 4 positive data points $\{1, 2, 3, 4\}$ and 3 negative data points $\{-3, -2, -1\}$. We want to learn a soft-margin SVM (though the dataset is linearly separable) for this dataset. Think those points on the real line.

Question 6

Statement

If $C \to \infty$, how many support vectors do we have?

Answer

2

Solution

When $C \to \infty$, the soft margin problem turns out to be the hard margin and it is clear that the dataset is linearly separable. Therefore, the decision boundary will be the origin and the support vectors will be -1 and 1. So, there will be two support vectors.

Question 7

Statement

If $C = 0$, how many support vectors do we have?

Answer

0

Solution

If $C = 0$, then the problem collapses as $w = 0$ and the concept of support vectors becomes moot. So, there will be no support vectors as for support vectors, $w^T x_i = \pm 1$ but here, $w = 0$.

Question 8

Statement

Is the following statement true or false?

For a fixed size of the training and test set, increasing the complexity of the model always leads to a reduction of the test error.

Options

(a)

True

(b)

False

Answer

(b)

Solution

If we increase the complexity of the model, the model tends to overfit as it tries to fit each and every data point including noise in the training dataset but may fail to classify the test data point. Therefore, the model will tend to overfit, and test errors may go up.

Question 9

Statement

If we remove all the non-support vectors from the dataset, what will be the effect on the model?

Options

(a)

Model will overfit

(b)

Model will underfit

(c)

The model will not be changed

(d)

Can not say

Answer

(c)

Solution

Remember that the weight vector in the SVM problem comes out to be

$$w^* = \sum_{i=1}^{n} \alpha_i^* x_i y_i$$

That is only the points for which $\alpha_i > 0$ (the support vectors) decide on the weight vector.

Therefore, if we remove the non-support vectors from the dataset, the model will not change.