**Birla Institute of Technology & Science, Pilani**

**Work-Integrated Learning Programmes Division**

**MTech. Software Engineering at DSE (FC04, FA04\_1-2021) Cluster**

**Second Semester 2021-2022**

**Endsem -Semester Test**

**(EC-3 Makeup)**

Course No. : DSECLZG565

Course Title : Machine Learning

No. of Pages = 3

No. of Questions = 8

Nature of Exam : Open Book

Weightage : 40%

Duration : 2 Hours

Date of Exam :

Note:

1. Please follow all the *Instructions to Candidates* given on the cover page of the answer book.
2. All parts of a question should be answered consecutively. Each answer should start from a fresh page.
3. Assumptions made if any, should be stated clearly at the beginning of your answer.

1. **.** Consider the following training set in 2-dimensional Euclidean space. [5 Marks}

|  |  |  |
| --- | --- | --- |
| Point | Coordinate | Class |
| X1 | (-1, 1) | Negative |
| X2 | (0, 1) | Positive |
| X3 | (0, 2) | Negative |
| X4 | (1, -1) | Negative |
| X5 | (1, 0) | Positive |
| X6 | (1, 2) | Positive |
| X7 | (2, 2) | Negative |
| X8 | (2, 3) | Positive |

What is the class of the point (1, 1) if 7NN classifier is considered? If the value of K is reduced whether the class will change? (Consider K=3 and K=5). What should be the final class if the above 3 values of K are considered?

## Solution: Euclidian distance following table – 2M

|  |  |  |  |
| --- | --- | --- | --- |
| Point | Coordinate | Class | Distance from 1,1 |
| X1 | (-1, 1) | Negative | 2 |
| X2 | (0, 1) | Positive | 1 |
| X3 | (0, 2) | Negative | 1.414 |
| X4 | (1, -1) | Negative | 2 |
| X5 | (1, 0) | Positive | 1 |
| X6 | (1, 2) | Positive | 1 |
| X7 | (2, 2) | Negative | 1.41 |
| X8 | (2, 3) | Positive | 2.236 |

1. class of the point (1, 1) if 3NN classifier is considered x2, x5, x6 - Positive
2. class of the point (1, 1) if 5NN classifier is considered?

x2, x5, x6, x3, x7 - Positive

1. class of the point (1, 1) if 7NN classifier is considered?

x2, x5, x6, x3, x7, x1, x5- Negative

Final class value to be considered as Positive

1. Use kernel trick and find the equation for hyperplane using nonlinear SVM. Positive

Points: {(7,0), (9,0), (11,0)} Negative Points: {(0,0), (8,0), (12,0), (10,0)}. Plot the point

before and after the transformation. [5 Marks]

Solution:

Φ(x) = x mod 2 [3M]

Equation of hyperplane : y=0.5 [2M]

1. Consider the following dataset. Fit a regression model (lines passing through the origin) used for line fitting for the given data, [4+3=7 Marks]

|  |  |
| --- | --- |
| x | y |
| 1 | 3 |
| 1 | 2 |
| 2 | 1 |
| 3 | 3 |
| 5 | 3 |
| 7 | 7 |
| 6 | 4 |
| 7 | 6 |
| 6 | 7 |
| 4 | 5 |

1. Compute the squared error E(w) for varying values of w ={-1.0, -0.2, -0.4, 0, 0.2, 0.4, 1.0}. Mark the point of global minima on the plotted error curve.
2. Define the best hypothesis and predict the value of x=7.6. Show the computation steps clearly.

Solution:

1. 2

At Global minima:

|  |  |
| --- | --- |
| W\* | e(w) |
| 0.92 | 15.5664 |

E(w) is a vertically oriented parabola with vertex (0.92, 15.56)

|  |  |
| --- | --- |
| w | E(w) |
| -1 | 849 |
| -0.2 | 299.24 |
| -0.4 | 409.56 |
| 0 | 207 |
| 0.2 | 132.84 |
| 0.4 | 76.76 |
| 1 | 17 |

1. Best hypothesis, and h(x=7.6) = 0.92 \* 7.6 =6.992=7
2. Consider the dataset given below where and are attributes which can take the values 0 and

1, and is the classification. The values marked “\*” represent data values that are corrupted. It

is known that during the construction of a decision tree to represent the clean dataset (i.e one

without any “\*”), the attribute was chosen at the root instead of attribute using information

gain. Is this information enough to guess the value of the bit that must replace “\*”? Give a detailed justification for your answer. [7 Marks]

|  |  |  |
| --- | --- | --- |
| A | B | Y |
| 1 | 0 | no |
| 1 | 1 | no |
| 0 | \* | no |
| 0 | 1 | yes |
| 0 | 1 | yes |
| 1 | 1 | yes |

**Answer**

**(2M)**

Let S be the given dataset. We have

.

(2M)

If we assume \*=1, then we

(2M)

If we assume \*=0, then we have

(2M)

Thus regardless of whether \*= 0 or 1, has a higher information gain than and would have been chosen to be the root. Thus the information given is not sufficient to decide the value of \*.

1. Let say we have a dice of 4 sides. Where

{ (x,P(x)) : (0,a), (1,(1-a)/3), (2,(1-a)/2), (3,(1-a)/6) }. If the tossing event is observed as

(0,1,2,3,2,3,1,0) then what is the most probable value of *a* [3 Marks]

Ans:

As each events are independent, the likelihood of the event would be

(P(x=0)\*P(x=1)\*P(x=2)……)

= a\*(1-a)/3\*(1-a)/2\*(1-a)/6\*(1-a)/2\*(1-a)/6\*(1-a)/3\*a

So we maximize (ignoring constants in denominator) = a^2\*(1-a)^6

Taking gradient and equal to zero

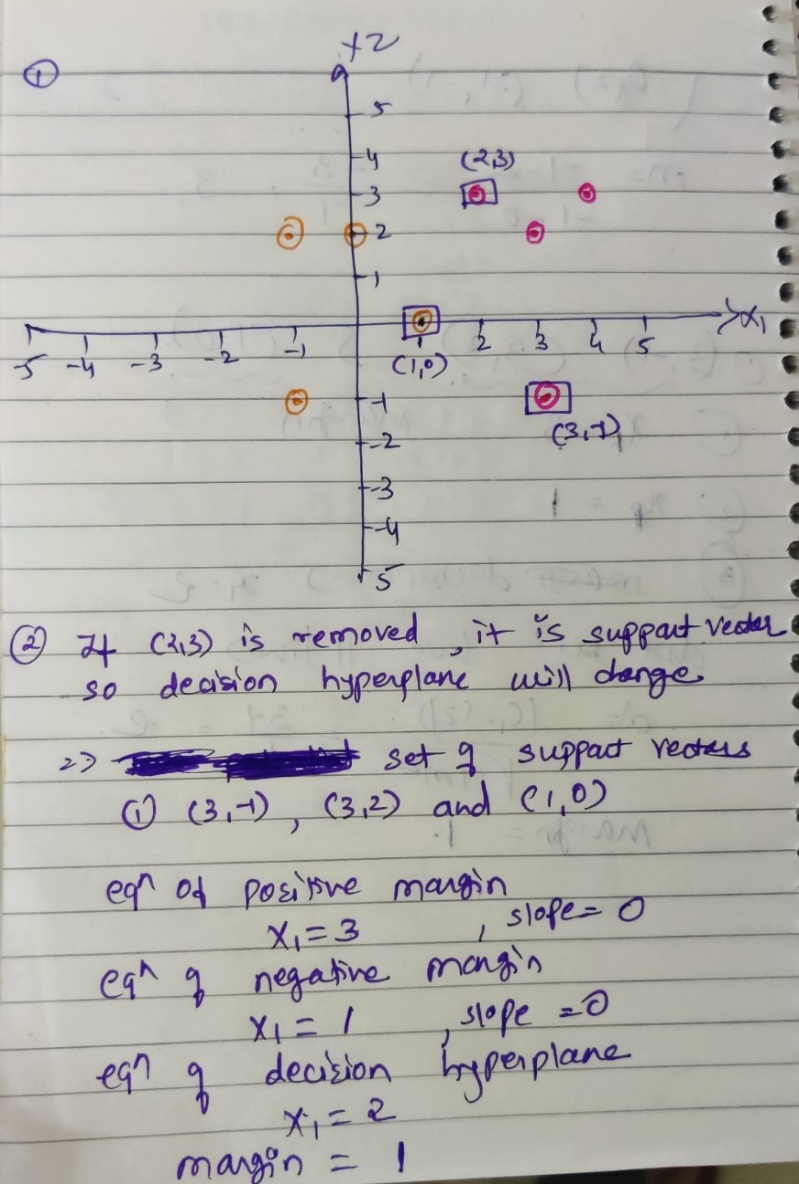
2\*a\*(1-a)^6 - 6\*a^2\*(1-a)^5=0

* (1-a)-3a =0
* a = ¼

1. For a linear Support Vector Machine method, positive Points are {(3, 2), (4, 3), (2, 3), (3, -1)} and Negative Points are{(1, 0), (-1, -1), (0, 2), (-1, 2)} [1+4=5Marks]
2. Find the support vectors
3. Determine the equation of hyperplane if it is changed and give a reason if it is not changed for the following two cases
   1. If the point (2, 3) is removed.
   2. If the point (-2,-3) is added

Solution:

1. Support vectors are (2,3), (1,0) and (3,-1) [1M- if one of the SVs is wrong then 0 M]
2. a. If the point (2, 3) is removed. [3M – 1M reason, 2 M for decision boundary equation]



The solution obtained using the Lagrange method is equally acceptable.

1. If the point (-2,-3) is added [1M – 0M if no reason is given]: The equation of the decision hyperplane will not change as the added point is not a support vector.
2. Consider training a boosting classifier using decision stumps on the following data set.

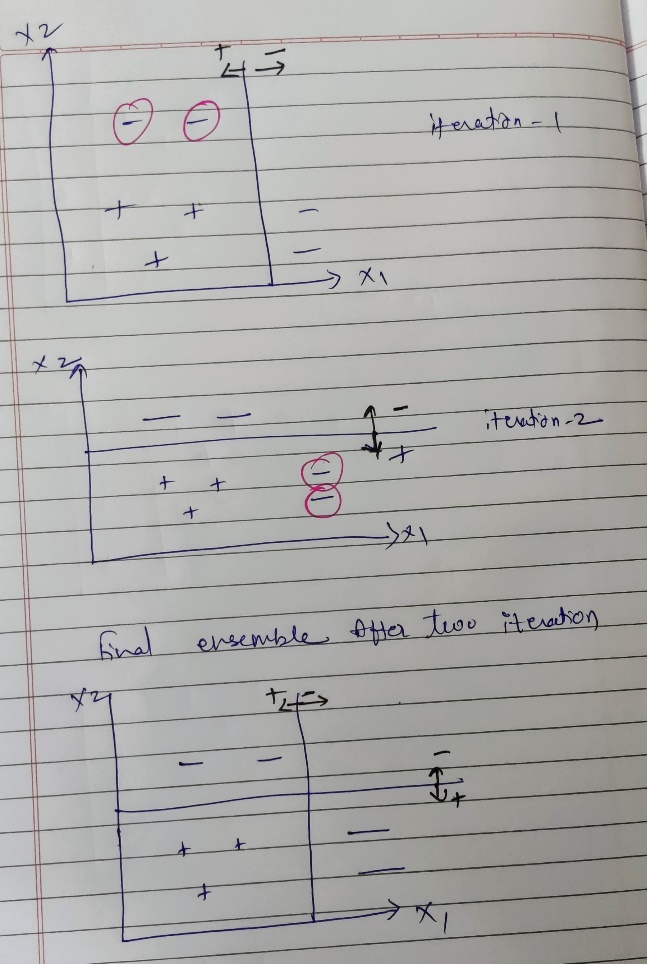
Circle the examples which will have their weights increased at the end of each

iteration. Run the iteration till zero training error is achieved. [3 Marks]

X1

X2

Solution: No of iteration - 2



1. Assume that a number of points are distributed along the x-axis:

and an outlier point at

We would like to use the -Means algorithm to find two clusters for these points.

Initially, one cluster center is chosen to be . Where should the other cluster center be placed on the x-axis initially so that one of the clusters formed has all the given data points and the other cluster has none?

What will be the final locations of the cluster centers? [2+3= 5 Marks]

**Answer**:

The second cluster center should be chosen initially to be where either or .

For these choices of the second center, the -Means algorithm will ensure that all the given data points are assigned to the first cluster center since every data point will be closer to the first cluster center than the second one.

In the second iteration of the algorithm the first cluster center will be updated from to and the second cluster center will be unchanged.