**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 Regular)**

Course No. : DSECLZG565

Course Title : Machine Learning

No. of Pages = 2

No. of Questions = 8

Nature of Exam : Open Book

Weightage : 40%

Duration : 2 Hours

Date of Exam : 25-09-22(FN)

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.
4. The results of the election are to be predicted for candidates based on dataset D.

There are three different hypotheses h1, h2 and h3 are used to predict the result

of candidates winning or losing an election. The probability of h1 given dataset D

is 0.5, the probability of h2 given dataset D is 0.3 and the probability of h3 given

dataset D is 0.2. Given a new candidate, h1 predicts that a candidate will win the

election whereas h2 and h3 predict that candidate will lose the election. What’s

the most probable classification of a new candidate? [3 Marks]

Solution:

+ = win, - = lose

*P*(*h*1|*D*) = .5, *P*(−|*h*1) = 0, *P*(+|*h*1) = 1

*P*(*h*2|*D*) = .3, *P*(−|*h*2) = 1, *P*(+|*h*2) = 0

*P*(*h*3|*D*) = .2, *P*(−|*h*3) = 1, *P*(+|*h*3) = 0



P(+|D) = 1\*0.5+0\*0.3+0\*0.2=0.5 **[1.5M]**

P(-|D) = 0\*0.5+1\*0.3+1\*0.2=0.5 **[1.5M]**

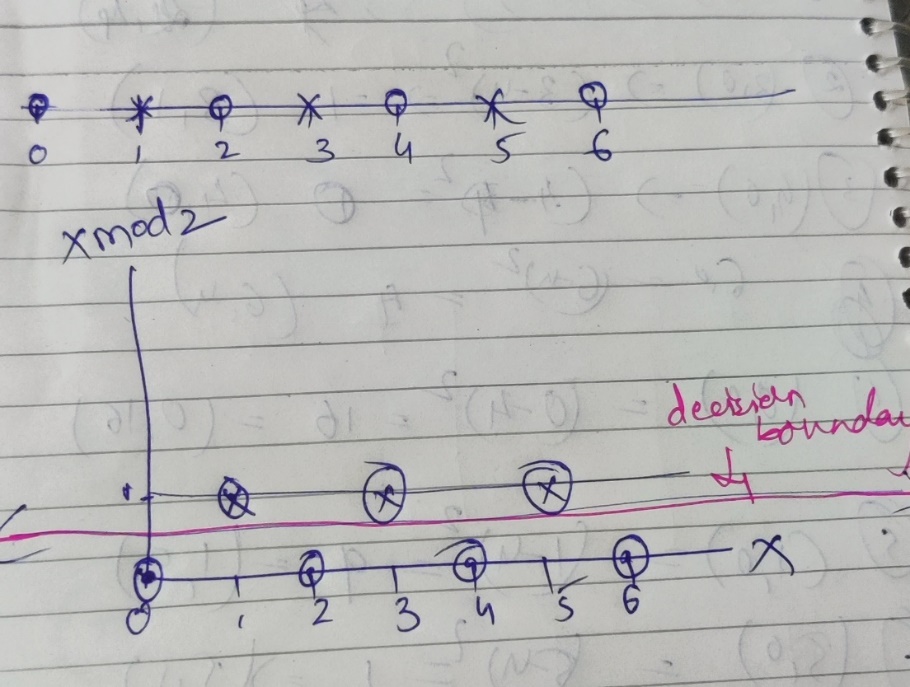
Winning and losing both are equiprobable

1. Use kernel trick and find the equation for hyperplane using nonlinear SVM. Positive Points: {(1,0), (3,0), (5,0)} Negative Points: {(0,0), (2,0), (4,0), (6,0)}. 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. Suppose we have the following one-dimensional data at -4.0, -.3.0, -2.0, -1.0, 0.0, 1.0, 2.0, 3.0, 4.0. Use the EM algorithm to find a Gaussian mixture model consisting of exactly one Gaussian that fits the data. Assume that the initial mean of the Gaussian is 10.0 and the initial variance is 1.0. [7 Marks]

**Answer**

First we note that since there is only one Gaussian in the mixture model.

Computing the posterior probabilities we see that the posterior probabilities are all equal to 1 since both the numerator and denominator are equal to . **[1.5 M]**

Also , the number of data points. [**0.5M]**

This completes the E-step.

In the M-step, we see that

**[2M]**

and .

Here the and are matrices and the expression for simplifies to

which is =6.66. [2M]

In the next iteration the E-step computes the posterior probabilities to be 1 and the M-step computes the same mean and covariance matrix as above, so the algorithm converges.[**1M]**

1. A dataset consists of the results of 100 independent coin tosses of the same

coin where 30 turn out to be heads and 70 turn out to be tails. Let be the

probability of tossing a head. How many datasets on 100 coin tosses are possible

which have the same likelihood as the given dataset ? Determine the maximum

likelihood estimate of the parameter using appropriate calculations. [5 Marks]

**Answer**: The likelihood of the given data given is .

Any other dataset with the same number of heads and tails as will have the same likelihood given the same probability of tossing a head. **[1M]**

There are ways of choosing locations out of to place heads.

Therefore the number of datasets that have the same likelihood as is . **[2M]**

To calculate the value of that maximizes likelihood we take log to get

Then taking the derivative of and setting it to zero,

we get .

Solving for we get . **[2M]**

1. Consider a following dataset [ 5 Marks]

|  |  |  |
| --- | --- | --- |
| X1 | X2 | Y |
| -1 | -1 | Positive class |
| -1 | 1 | Negative class |
| 1 | -1 | Negative class |
| 1 | 1 | Positive class |

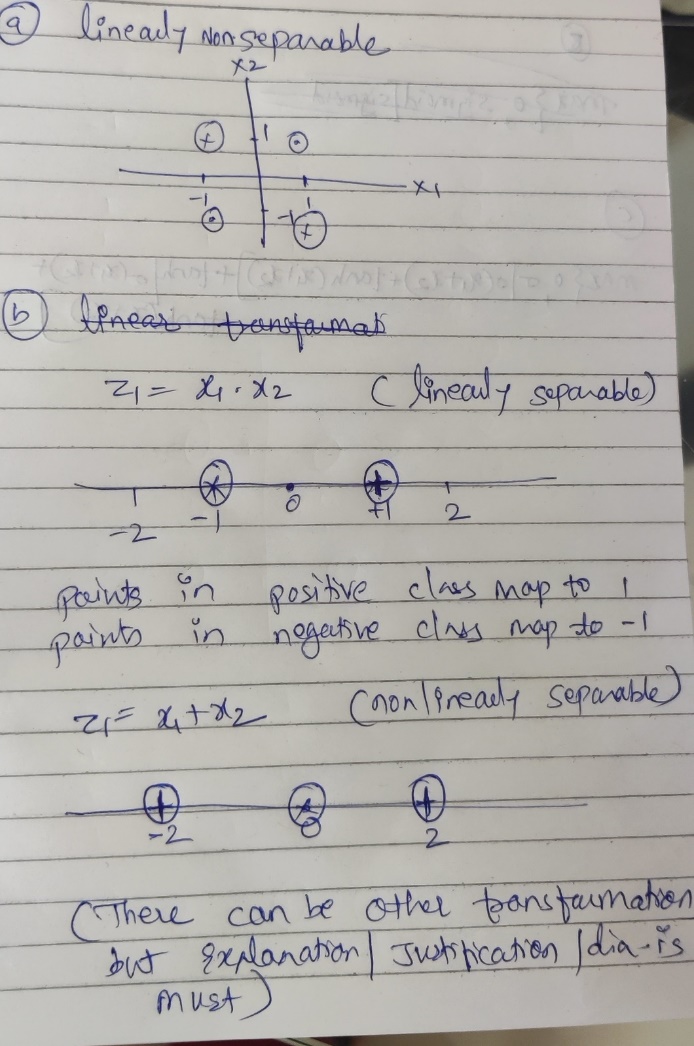
Answer the following with respect to above dataset

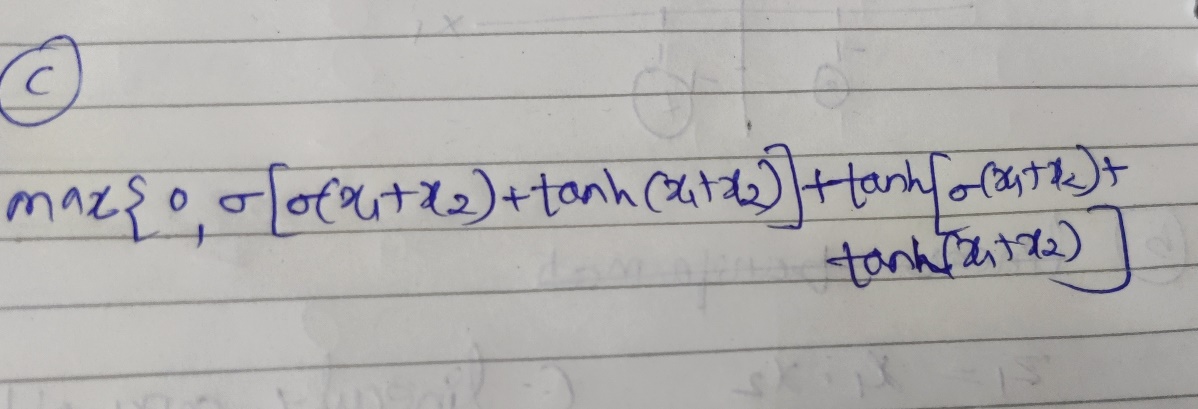
1. Comment on the separability of the dataset with an explanation.[0.5M]
2. Provide the 1-dimensional transformation of this dataset for each linearly and non-linearly separable case with justification[2M]
3. Model the above dataset with an Artificial Neural Network which has two hidden layers, each of which contains two units. Assume that the weights in each layer are set to 1 so that the top unit in each layer applies sigmoid activation to the sum of its inputs and the bottom unit in each layer applies tanh activation to the sum of its inputs. Finally, the single output node applies ReLU activation to the sum of its two inputs. Write the output of this neural network in closed form as a function of x1 and x2. (no need to calculate exact values) [2.5M]

Solution:

In b) just transformation with no explanation deduct 0.5M in each case.

In c) any mistake in function – 0 M





1. Solve the below and find the equation for hyper plane using linear Support Vector

Machine method. Positive Points: {(3, 2), (4, 3), (2, 3), (3, -1)} Negative Points: {(1, 0),

(-1, -1), (0, 2), (-1, 2)} [5 Marks]

1. Find the support vectors
2. 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 (5,4) 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]

Equation of decision Hyperplane - 2M

The solution obtained using the Lagrange method or by geometrical inspection is acceptable

1. a. If the point (2, 3) is removed. [1M – 0.5M if no reason is given]:

The equation of the decision hyperplane will change as the removed point is the support vector.

1. If the point (5,4) is added [1M – 0.5M if no reason is given]: The equation of the decision hyperplane will not change as the added point is not 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]



Solution:

No of iteration - 3



1. Consider the following dataset. [7 Marks]

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| x1 | -1 | -1 | -1 | -1 | 0 | 4 | 4 |
| x2 | 2 | 1 | -1 | -2 | 0 | 2 | -2 |
| Class label | 1 | 1 | 1 | 1 | 1 | 2 | 2 |

The given class label exhibits two natural clusters formed in the given dataset and acts as a ground truth. Now remove class labels and use the K-means clustering algorithm to find the 2 clusters by initializing two cluster centres as follows:

1. C1(-1,2) and C2(0,0)
2. C1(-0.5,0) and C2(0,0)

For both the above cases run the algorithm till centres do not change (convergence criteria) and give the final cluster assignment [2+2M]. In each case, comment on the correctness of cluster assignment. [1+1M] Also, comment in no more than 20 words on the drawback of k-means which is depicted in above two cases.[1M]

Solution:

A.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | c1 |  |  |  | c2 |  |  |
| x1 | -1 | -1 | -1 | -1 | 0 | 4 | 4 |
| x2 | 2 | 1 | -1 | -2 | 0 | 2 | -2 |
| d1 | 0 | 1 | 3 | 4 | 2.236068 | 5 | 6.403124 |
| d2 | 2.236068 | 1.414214 | 1.414214 | 2.236068 | 0 | 4.472136 | 4.472136 |
| cluster | 1 | 1 | 2 | 2 | 2 | 2 | 2 |
|  |  |  |  |  |  |  |  |
|  | new c1 | new c2 |  |  |  |  |  |
| x1 | -1 | 1.2 |  |  |  |  |  |
| x2 | 1.5 | -0.6 |  |  |  |  |  |
| d1 | 0.5 | 0.5 | 2.5 | 3.5 | 1.802776 | 5.024938 | 6.103278 |
| d2 | 3.405877 | 2.720294 | 2.236068 | 2.607681 | 1.341641 | 3.820995 | 3.130495 |
| cluster | 1 | 1 | 2 | 2 | 2 | 2 | 2 |
|  |  |  |  |  |  |  |  |
|  | new c1 | new c2 |  |  |  |  |  |
| x1 | -1 | 1.2 |  |  |  |  |  |
| x2 | 1.5 | -0.6 |  |  |  |  |  |

Comment on cluster assignment:

Algorithm has converged after 2 iterations but the cluster assignment does not depict the natural clusters in the datasets as given by the ground truth.

B

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | c1 |  |  |  | c2 |  |  | c1 |
| x1 | -1 | -1 | -1 | -1 | 0 | 4 | 4 | -0.5 |
| x2 | 2 | 1 | -1 | -2 | 0 | 2 | -2 | 0 |
| d1 | 2.061553 | 1.118034 | 1.118034 | 2.061553 | 0.5 | 4.924429 | 4.924429 |  |
| d2 | 2.236068 | 1.414214 | 1.414214 | 2.236068 | 0 | 4.472136 | 4.472136 |  |
| cluster | 1 | 1 | 1 | 1 | 2 | 2 | 2 |  |
|  |  |  |  |  |  |  |  |  |
|  | new c1 | new c2 |  |  |  |  |  |  |
| x1 | -1 | 2.666667 |  |  |  |  |  |  |
| x2 | 0 | 0 |  |  |  |  |  |  |
| d1 | 2 | 1 | 1 | 2 | 1 | 5.385165 | 5.385165 |  |
| d2 | 4.176655 | 3.800585 | 3.800585 | 4.176655 | 2.666667 | 2.403701 | 2.403701 |  |
| cluster | 1 | 1 | 1 | 1 | 1 | 2 | 2 |  |
|  |  |  |  |  |  |  |  |  |
|  | new c1-mean | new c2 |  |  |  |  |  |  |
| x1 | -0.8 | 4 |  |  |  |  |  |  |
| x2 | 0 | 0 |  |  |  |  |  |  |
| d1 | 2.009975 | 1.019804 | 1.019804 | 2.009975 | 0.8 | 5.2 | 5.2 |  |
| d2 | 5.385165 | 5.09902 | 5.09902 | 5.385165 | 4 | 2 | 2 |  |
| cluster | 1 | 1 | 1 | 1 | 1 | 2 | 2 |  |
|  | new c1 | new c2 |  |  |  |  |  |  |
| x1 | -0.8 | 4 |  |  |  |  |  |  |
| x2 | 0 | 0 |  |  |  |  |  |  |

Comment on cluster assignment:

Algorithm has converged after 3 iterations and the cluster assignment shows the natural clusters in the datasets as given by the ground truth.

The drawback of k-means as demonstrated by the above two cases:

Correctness of the K- means the algorithm is sensitive to the initialization of cluster centres.