G.PULLAIAH COLLEGE OF ENGINEERING AND TECHNOLOGY[AUTONOMUS]
III B.Tech II Semester (RU-20) I-Assignment Submission March 2023
COMPUTER SCIENCE AND ENGINEERING
Name of the Student : Roll No. : Devathi.Srija, 20ATIA05F5
COURSE TITLE : Machine Learning (Course Code: A30528

1. Describe the Find-S algorithm and candidate elimination algorithm, Explain by taking Enjoysport concept and training instance given below.

Example	Sky	AirTemp	Humidity	Wind	Water	Forecast	Enjoy sport
1.	sunny	warm	Normal	strong	warm	same	yes
2.	sunny	warm	high	strong	warm	same	yes
3.	Rainy	cold	high	strong	warm	change	no
4.	sunny	warm	high	strong	cool	change	yes

A. Find-S algorithm:
→ This algorithm considers only positive examples
→ Most specific hypothesis => ϕ
→ Most general hypothesis => ?

Algorithm:
step1: Initialize h with most specific hypothesis (ϕ) in H
$h_0 = \langle \phi, \phi, \phi, \phi, \phi \rangle$ => 5 attributes (depends on attributes)

step2: For each positive sample,
For each attribute,
if (value = hypothesis value) => ignore
else
Replace with the most general hypothesis (?)

step3: output hypothesis h

Example:
Applying Find-S algorithm on given Example Enjoysport

step1: $h_0 = \langle \phi, \phi, \phi, \phi, \phi, \phi \rangle$
step2: $h_1 = \langle$sunny, warm, normal, strong, warm, same$\rangle$
step3: $h_2 = \langle$sunny, warm, ?, strong, warm, same$\rangle$
step4: $h_3 = h_2$
step5: $h_4 = \langle$sunny, warm, ?, strong, ?, ?$\rangle$

Disadvantage:
→ considers only positive (yes) values
→ h_4 may not be sole hypotheses that fits the complete data.

Candidate elimination Algorithm:

→ It uses the concept of version space
→ It consider both positive and negative values (Yes & No)
→ Both specific and general hypothesis
→ For positive samples, move from specific to general
→ For negative samples, move from general to specific
→ specific values $<\phi, \phi, \phi, \phi, \phi>$ → positive
→ General uses $<?, ?, ?, ?, ?>$ → Negative

Algorithm:

Step 1: Initialize both general and specific hypothesis (S and G)

$S = \{\phi, \phi, \phi, \phi, \phi\}$
$G = \{?, ?, ?, ?, ?\}$ depends on number of attributes

Step 2: For each example,
if example is positive
Make specific to general
if example is negative
Make general to specific

Example:

Applying candidate elimination algorithm on given example enjoy sport.

$S0 = \{\phi, \phi, \phi, \phi, \phi, \phi\}$ $G0 = \{?, ?, ?, ?, ?, ?\}$

1. For positive : $S1 = \{$'sunny', 'warm', 'normal', 'strong', 'warm', 'same'$\}$

$G1 = \{?, ?, ?, ?, ?, ?\}$

2. For positive : $S2 = \{$'sunny', 'warm', ?, 'strong', 'warm', 'same'$\}$

$G2 = \{?, ?, ?, ?, ?, ?\}$

3. For negative : $S3 = \{$'sunny', 'warm', ?, 'strong', 'warm', 'same'$\}$

$G3 = \{<$'sunny', ?, ?, ?, ?, ?$>, <?,$ 'warm', ?, ?, ?, ?$>, <?, ?, ?, ?, ?,$ 'same'$>\}$

4. For positive : $S4 = \{$'sunny', 'warm', ?, 'strong', ?, ?$\}$

$G4 = \{<$'sunny', ?, ?, ?, ?, ?$>, <?,$ 'warm', ?, ?, ?, ?$>\}$

Therefore S4 and G4 are final hypothesis

G PULLAIAH COLLEGE OF ENGINEERING AND TECHNOLOGY[AUTONOMUS]
III B Tech II Semester (RU 20) I Assignment Submission March 2023
COMPUTER SCIENCE AND ENGINEERING
Name of the Student Roll No Bevathi Sriya 20CA1A0515
COURSE TITLE Machine Learning (Course Code 20CS623)

2. Give Decision tree for the following set of training examples. Calculate entropy, information gain and gini index.

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	weak	NO
D2	Sunny	Hot	High	strong	NO
D3	Overcast	Hot	High	weak	yes
D4	Rain	mild	High	weak	yes
D5	Rain	cool	Normal	weak	yes
D6	Rain	cool	Normal	strong	NO
D7	Overcast	cool	Normal	strong	yes
D8	Sunny	mild	High	weak	NO
D9	Sunny	cool	Normal	weak	yes
D10	rain	mild	Normal	weak	yes
D11	Sunny	mild	Normal	strong	yes
D12	Overcast	mild	High	strong	yes
D13	Overcast	hot	Normal	weak	yes
D14	rain	mild	High	strong	NO

Step1: Create a root Node

→ The attribute that best classifies the training data, use this attribute at the root of the tree

→ calculate entropy

$$\text{entropy} = \frac{-p}{p+n}\log_2\left(\frac{p}{p+n}\right) - \frac{n}{p+n}\log_2\left(\frac{n}{n+p}\right)$$

→ calculate average information

$$I(\text{attribute}) = \sum \frac{p_i+n_i}{p+n}\,\text{entropy}(A)$$

→ calculate Information gain

$$\text{Gain} = \text{entropy}(S) - I(\text{attribute})$$

Building decision tree for a given example

P = 9 N = 5 Total = 14

$$\text{Entropy} = \frac{-p}{p+n}\log_2\left(\frac{p}{p+n}\right) - \frac{n}{p+n}\log_2\left(\frac{n}{p+n}\right)$$

$$\text{entropy}(S) = \frac{-9}{9+5}\log_2\left(\frac{9}{9+5}\right) - \frac{5}{9+5}\log_2\left(\frac{5}{9+5}\right)$$

$$\text{Entropy}(S) = 0.940$$

→ For each attribute: (let say outlook)
calculate entropy for each values i.e for 'Sunny', 'Rainy', 'overcast'

Outlook	p	n	Entropy
Sunny	2	3	0.971
Rainy	3	2	0.971
overcast	4	0	0

calculate entropy (outlook = 'value'):

$$E(\text{Outlook} = \text{Sunny}) = \frac{-2}{5}\log\left(\frac{2}{5}\right) - \frac{3}{5}\log\left(\frac{3}{5}\right) = 0.971$$

$$E(\text{Outlook} = \text{overcast}) = -1\log(1) - 0\log(0) = 0$$

$$E(\text{Outlook} = \text{rainy}) = \frac{-3}{5}\log\left(\frac{3}{5}\right) - \frac{2}{5}\log\left(\frac{2}{5}\right) = 0.971$$

→ calculate Average Information Entropy

$$I(\text{Outlook}) = \frac{p_{\text{sunny}} + n_{\text{sunny}}}{p+n}\text{entropy}(\text{Outlook} = \text{Sunny}) + \frac{p_{\text{rainy}} + n_{\text{rainy}}}{p+n}\text{entropy}(\text{Outlook} = \text{rainy}) + \frac{p_{\text{overcast}} + n_{\text{overcast}}}{p+n}\text{entropy}(\text{Outlook} = \text{overcast})$$

$$I(\text{Outlook}) = \frac{3+2}{9+5} * 0.971 + \frac{2+3}{9+5} * 0.971 + \frac{4+0}{9+5} * 0 = 0.693$$

→ calculate Gain: attribute is outlook

$$\text{Gain}(\text{Outlook}) = 0.940 - 0.693 = 0.247$$

→ For each attribute (let say temperature)

Temperature	p	n	Entropy
Hot	2	2	1
Mild	4	2	0.918
Cool	3	1	0.811

COURSE TITLE: Machine Learning

→ calculate average information entropy

I(Temperature) = $\frac{2+2}{4+5}$ * 1 + $\frac{3+2}{4+5}$ * 0.918 + $\frac{2+1}{4+5}$ * 0.918

→ calculate Gain attribute is Temperature

Gain(Temperature) : 0.940 - 0.911 = 0.029

→ for each attribute : (Let say Humidity)

Humidity	P	n	Entropy
High	3	4	0.985
Normal	6	1	0.591

→ calculate average information entropy

I(Humidity) : $\frac{3+4}{9+5}$ * 0.985 + $\frac{6+1}{9+5}$ * 0.591 = 0.788

→ calculate gain : attribute is Humidity

Gain(Humidity) : 0.940 - 0.788 = 0.152

→ for each attribute : let say windy

Windy	P	n	Entropy
Strong	3	3	1
Weak	6	2	0.811

→ calculate Average Information entropy

I(windy) = $\frac{3+3}{9+5}$ * 1 + $\frac{6+2}{9+5}$ * 0.811 = 0.892

→ calculate Gain : attribute is windy

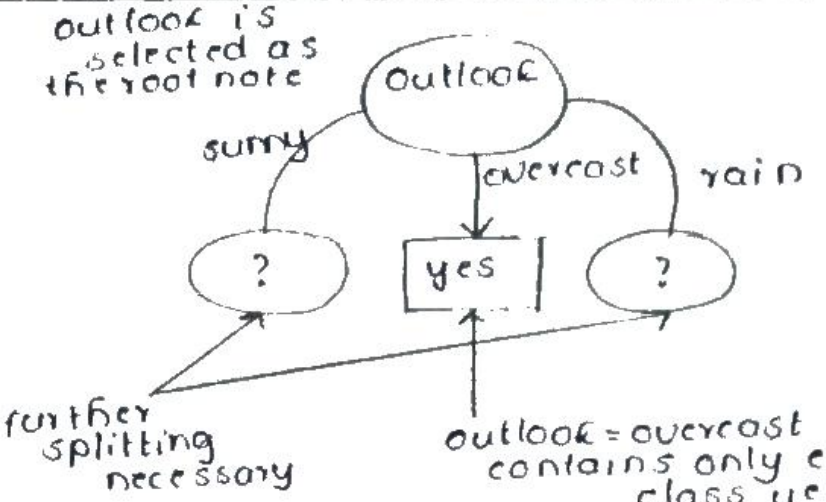
Gain(windy) : 0.940 - 0.892 = 0.048

→ Pick the highest gain attribute

Attributes	Gain
Outlook	0.247
Temperature	0.029
Humidity	0.152
windy	0.048

RESPONSE : Outlook

Outlook	Temperature	Humidity	Windy	Play Tennis
overcast	hot	high	weak	yes
overcast	cool	normal	strong	yes
overcast	mild	high	strong	yes
overcast	hot	normal	weak	yes



→ Repeat the same thing for subtrees till we get the tree

Outlook = sunny

p = 2, N = 3, Total = 5

$$\text{entropy}(S_{sunny}) = \frac{-2}{2+3}\log_2\left(\frac{2}{2+3}\right) - \frac{3}{2+3}\log_2\left(\frac{3}{2+3}\right) = 0.971$$

→ For each attribute: (Let say humidity)

→ calculate entropy for each humidity, i.e for 'high', 'Normal'

Humidity	p	n	Entropy
high	0	3	0
normal	2	0	0

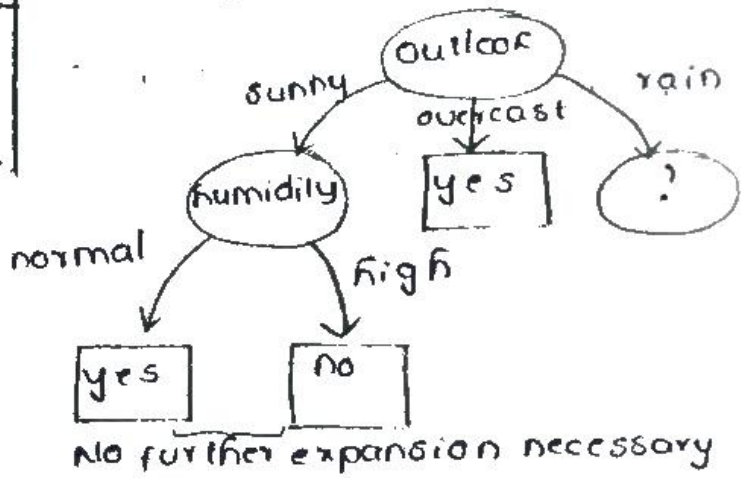
$I(\text{humidity}) = 0$

Gain = 0.971

→ Pick the heighest gain attribute

Attributes	Gain
Temperature	0.571
humidity	0.971
windy	0.02

Next Node: in sunny - humidity



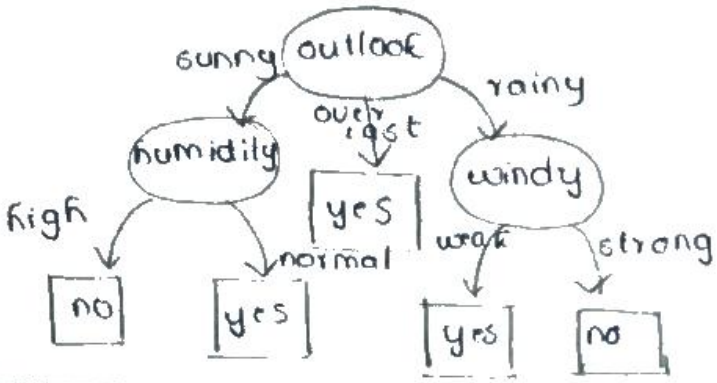
G.PULLAIAH COLLEGE OF ENGINEERING AND TECHNOLOGY[AUTONOMUS]
III B.Tech II Semester (RU-20) I-Assignment Submission March 2023
COMPUTER SCIENCE AND ENGINEERING
COURSE TITLE : Machine Learning Name of the Student : Roll No. : Devathi Grija, 20ATIA0515
(Course Code: A30528

→ Repeat the same process and the gain we get is

Attributes	Gain
Humidity	0.02
windy	0.971
Temperature	0.02

Next Node in Rainy : windy

Final decision tree



3 With a neat sketch explain how the two hyper planes in SUPPORT VECTOR MACHINE works.

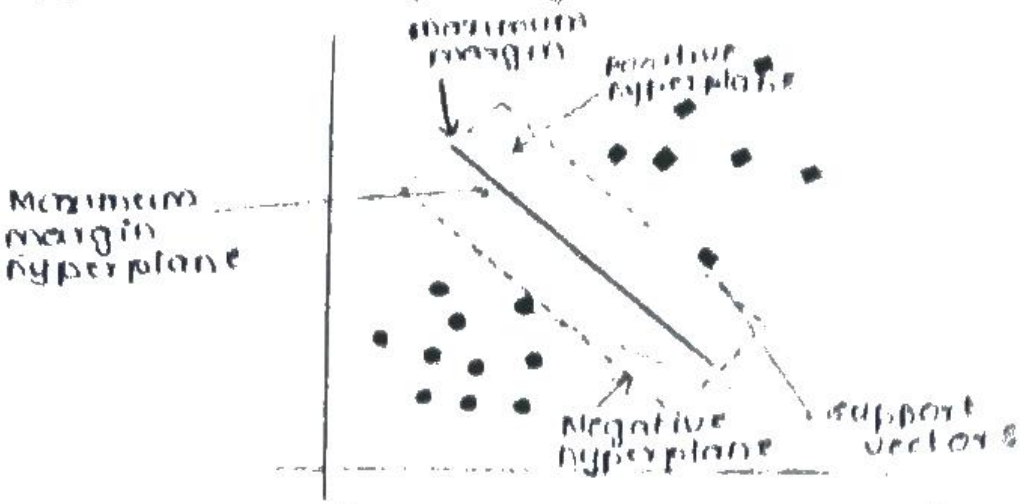
Support vector machine:

Support vector machine is one of the most popular supervised learning algorithms, which is used for classification as well as regression problem. However, it is used for classifica problems in machine learning.

The goal of SVM algorithm is to create the best line that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called hyper plane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors and hence algorithm is termed as support vector machine.

consider below diagram in which there are two different categories that are classified using a decision boundary or hyperplane



Hyperplane and support vectors in the SVM algorithm:

Hyperplane:

There can be multiple lines/decision boundaries to segregate the classes in n-dimensional space, but we need to find out the best decision boundary that helps to classify the data points. This best boundary is known as the hyperplane of SVM.

→ The dimensions of hyperplane depend on the features present in the dataset, which means if there are 2 features, then hyperplane will be a 2-dimension plane.

→ we always create a hyperplane that has a maximum margin, which means the maximum distance between the data points.

Support vectors:

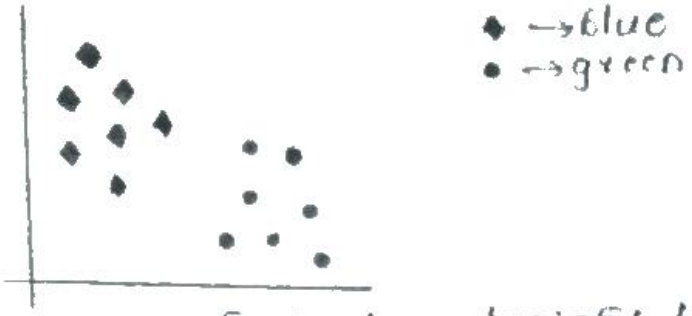
The data points or vectors that are the closest to the hyperplane and which affect the position of the hyperplane are termed as support vector since these vectors support the hyperplane, hence called a support vector.

How does SVM works:

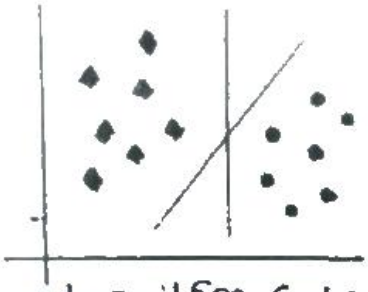
The working of the SVM algorithm can be understood by using an example. Suppose we have a dataset that has two tags (green and blue), and the dataset has two features x_1 and x_2. we want

G PULLAIAH COLLEGE OF ENGINEERING AND TECHNOLOGY[AUTONOMUS]
III B.Tech II Semester (RU-20) I-Assignment Submission March 2023
COMPUTER SCIENCE AND ENGINEERING
Name of the Student : Roll No : Revathi Priya, 20AT1A05F5
COURSE TITLE - Machine Learning (Course Code A30528

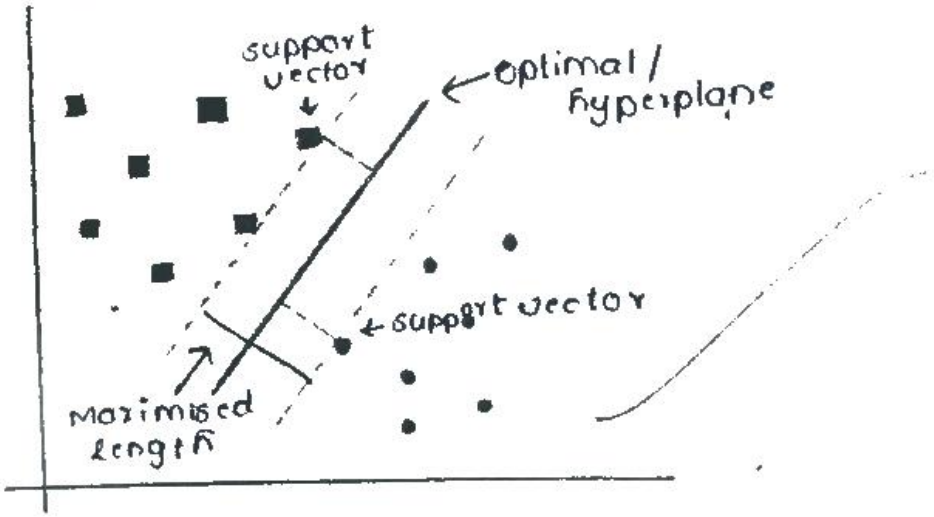
classifier that can classify the pair (x_1, x_2) of coordinates in either green or blue. consider the below image.



so as it is ad-space so by just a straight line. we can easily seperate these two classes. But there can be multiple lines that can separate these classes. consider the below image



Hence, the SVM algorithm helps to find the best line or decision boundary, this best boundary or region is called as a hyperplane. SVM algorithm finds the closest point of the lines from both the classes. These points are called support vectors. The distance between the vectors and the hyperplane is called as margin. And the goal of SVM is to maximize this margin. The hyperplane with maximum margin is called the optimal hyperplane.



4. Enumerate indetail how Ada Boost algorithm works.

AdaBoost:

Boosting is an ensemble modeling technique that attempts to build a strong classifier from the number of weak classifiers. It is done by building a model by using weak models in series. Firstly, a model is build which tries to build from training data. The the second model is built which tries to correct the errors present in the first model. This procedure is continued and models are added until either the complete training data set is predicted correctly or the maximum number of models are added.

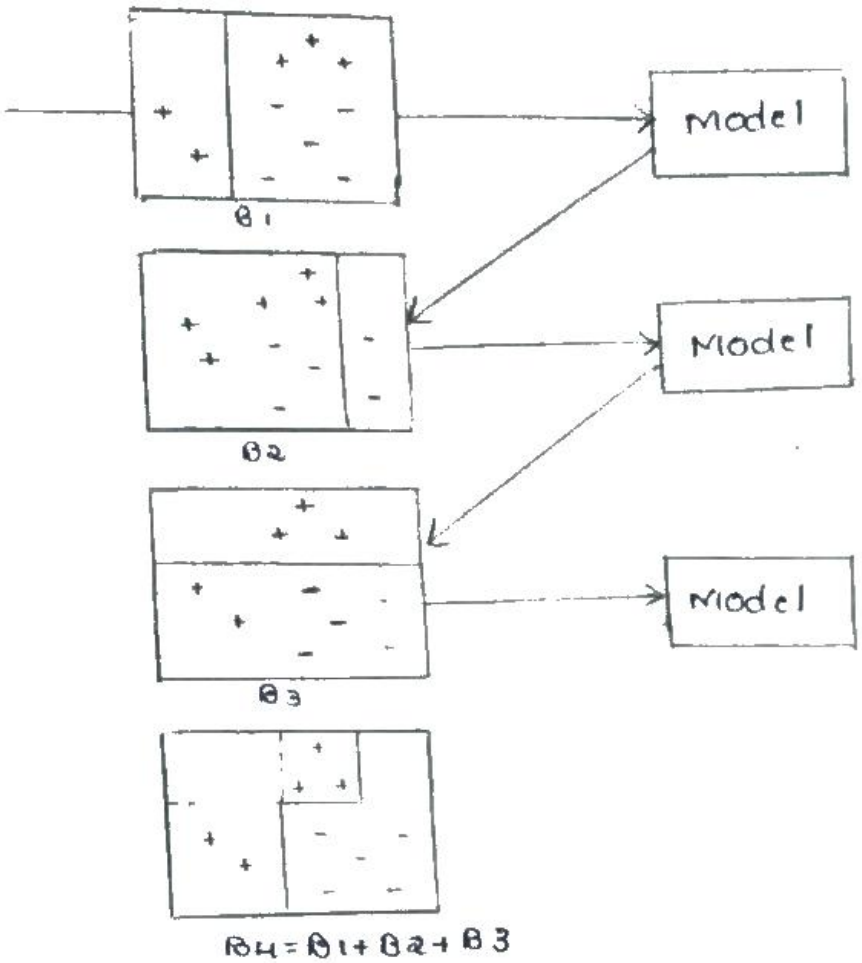
Ada Boost was the first really successful boosting algorithm developed for the purpose of binary classification. AdaBoost is short for Adaptive Boosting and is a very popular boosting technique that combines multiple "weak classifiers" into a single "strong classifier". It was formulated by Yoav Freund and Robert Schapir. They also won the 2003 Godel Prize for their work.

Algorithm:

1. Initialise the dataset and assign equal weight to each of the data point.
2. provide this as input to the model and identify the wrongly classified data points
3. Increase the weight of the wrongly classified data points
4. if(got required results)
   goto step5
   else
   goto step2

G. PULLAIAH COLLEGE OF ENGINEERING AND TECHNOLOGY[AUTONOMOUS]
III B Tech II Semester (RU-20) I-Assignment Submission March 2023
COMPUTER SCIENCE AND ENGINEERING
Name of the Student : Devathi. Sruja Roll No. : 20A91A05F5
COURSE TITLE : Machine Learning Course Code: A30528

Example:



Explanation:
The above diagram explains the AdaBoost algorithm in a very simple way. Let's try to understand it in a stepwise process:

→ B1 consists of 10 data points which consist of two types namely plus(+) and minus(-) and 5 of which are plus(+) and the other 5 are minus(-) and each one has been assigned equal weight initially. The first model tries to classify 3 plus(+) as minus(-)

→ B2 consists of the 10 data points from the previous model in which the 3 wrongly classified plus(+) are weighted more so that the current model tries more to classify these pluses(+) correctly. This model generates a vertical separator line that correctly classifies the

previously wrongly classified pluses(+) but in this attempt, it wrongly classifies three minuses(-).

→ B3 consists of the 10 data points from the previous model in which the 3 wrongly classified minus(-) are weighted more so that the current model tries more to classify these minuses(-) correctly. This model generates a horizontal seperator line that correctly classifies the previously wrongly classified minuses(-).

→ Bu combines together B1, B2 and B3 in order to build a strong prediction model which is much better than any individual model used.

Weight

21/3/2023