Pattern Recognition, Neural Networks and Deep Learning: Support Vector Machines (SVMs)

Rakan Zabian - 1741706 - MSc Data Science

05/04/2021

1 Write down the first 7 digits of your student ID as s1s2s3s4s5s6s7.

1741706

2 Find R1 which is the remainder of 4. Table 1 shows the multi-class methods to be used corresponding to the value of R1 obtained.

$\frac{1+7+4+1+7+0+6}{4} = 24R2$

R1 = 2
Method = Binary decision tree

3 Create a linearly separable two-dimensional dataset of your own, which consists of 3 classes. List the dataset in the format as shown in Table 2. Each class should contain at least 10 samples and all three classes have the same number of samples.

Sample of Class 1	Sample of Class 2	Sample of Class 3
(-1.131999 , 1.88043765)	(-1.99263752, -1.34550373)	(0.71496745, -0.86468779)
(-2.03086376, 2.9593066)	(-1.33470351, -2.31313259)	(0.85898249, -1.02307434)
(-2.42153118, 1.90683209)	(-0.76385808, -0.85498959)	(2.07882744, -2.06278951)
(-1.38749784, 2.44248049)	(-0.86459086, -3.3523126)	(2.7455513 , -2.66938057)
(-2.78950645, 1.79779494)	(-1.43007191, -3.62887262)	(3.34285577, -3.17667264)
(-3.22162231, 1.43168155)	(-2.42571434, -2.02017234)	(0.71393097, -0.83816387)
(-2.61467918, 1.67320028)	(-2.20843365, -4.32480657)	(1.36270369, -1.43092494)
(-2.09317588, 2.16194086)	(-2.02275279, -0.71508652)	(3.10696146, -3.03836399)
(-2.23094784, 2.42395791)	(-2.50791811, -2.41488504)	(2.65402605, -2.56514088)
(-2.34445659, 1.73207554)	(-1.7743148 , -1.67069488)	(1.97583188, -1.98774944)

Table 1: Samples of three classes.

This dataset of 10 samples per class was generated using 'sklearn.datasets.make_classification'. The parameter 'random_state' was equated to the student ID (1741706) to avoid any plagiarism/collusion issues.

4 Plot the dataset in Q3 to show that the samples are linearly separable. Explain why your dataset is linearly separable.

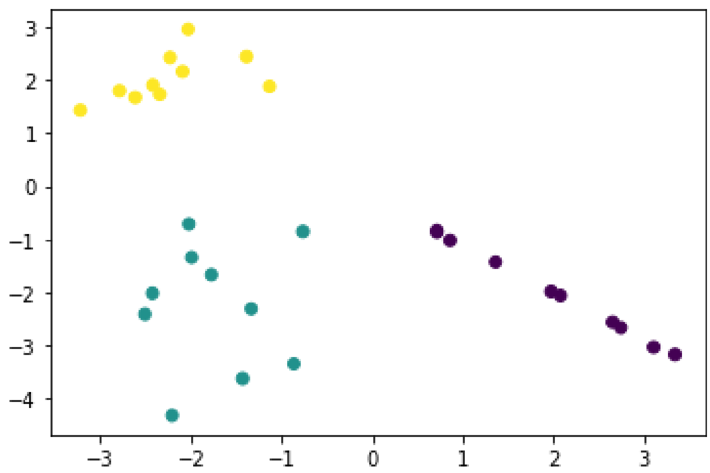

Figure 1: Scattered dataset

By inspection of figure 1, we observe that the most optimal way (largest margin) to separate the above classes linearly using the Binary decision tree approach is to separate class 1 (Yellow) from classes 2 (Grey) and 3 (Navy), and to then separate 2 (G) and 3(N) from one another.

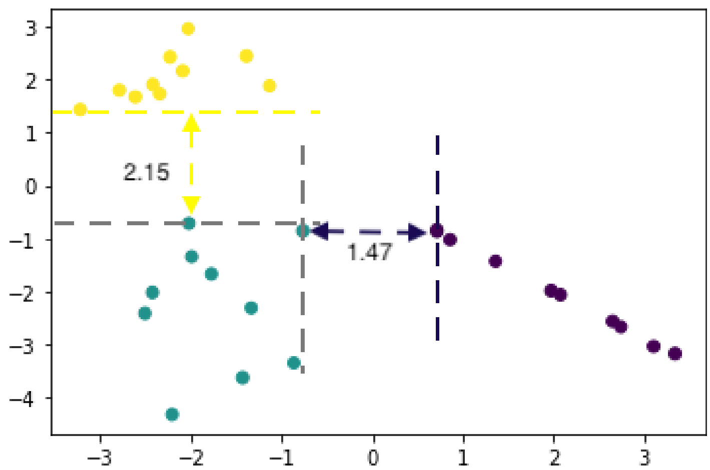

Figure 2: Annotated margins

In figure 2, we see that the distance between the lowest point in class 1 (Y) and the highest point in class 2 (G) is larger than the distance between the rightmost point in class 2 (G) and the leftmost point in class 3 (N). Hence, the assumption made by the inspection is justified as the largest margins separating the classes are displayed.

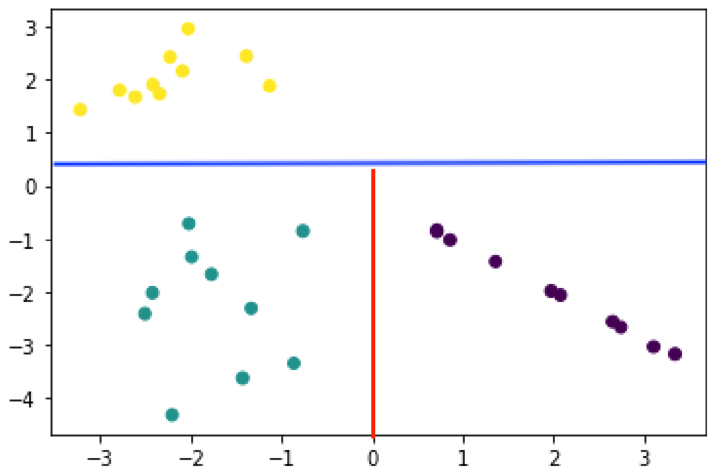

Figure 3: Linear separability

In figure 3, we observe that the three classes can be separated using straight lines. Hence, the dataset is linearly separable. The blue line represents the hyperplane that linearly separates class 1 (Y) from classes 2 (G) and 3 (N). The red line represents the hyperplane that linearly separates class 2 (G) from class 3 (N).

5 According to the method obtained in Q2, draw a block diagram at SVM level to show the structure of the multi-class classifier constructed by linear SVMs. Explain the design (e.g., number of inputs, number of outputs, number f SVMs used, class label assignment, etc.) and describe how this multi-class classifier works.

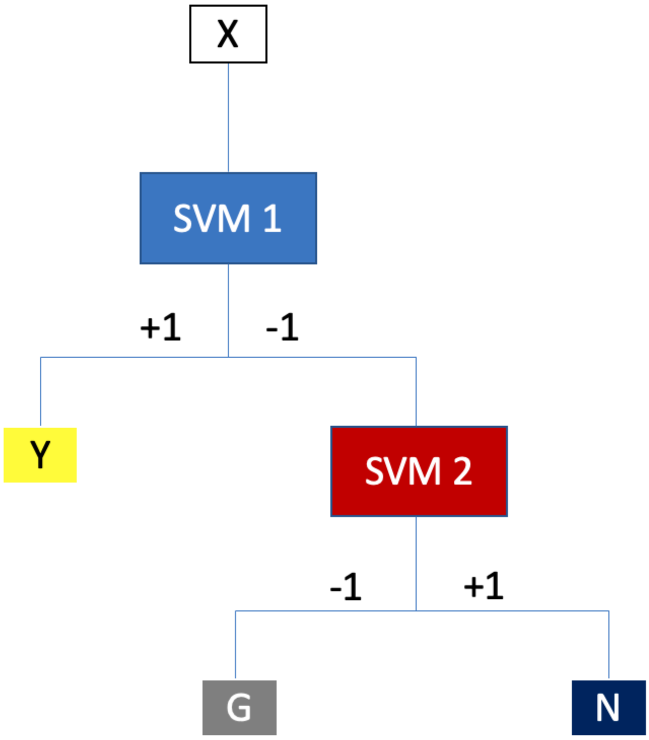

Figure 4: Block diagram

In figure 4, we observe a structure explaining how two SVMs combine to form a multi-class classifier using the Binary decision tree approach. The input, sample X, is classified by SVM 1, represented by the blue line in figure 3. SVM 1 will output a class label +1 or -1, classifying X as either a sample belonging to class 1 (Y), or as a sample that may belong to any of classes 2 (G) or 3 (N). If the output of SVM 1 is +1, X will be classified as class 1 (Y). If the output is -1, X will be input into SVM 2, represented by the red line in figure 3. SVM 2 will output a class label +1 or -1, classifying X as either a sample belonging to class 2 (G), or as a sample belonging to class 3 (N).

Inputs: As shown in figure 4, the input is a sample X, which is a two dimensional feature vector. Hence, there are two inputs to each SVM.

Outputs: As shown in figure 4, the output of each SVM is a class label (+1 or -1). Hence, there is one output for each SVM.

Class label assignment: We define the label for class 1 (Y) as +1, class 2 (G) as -1, and class 3 (N) as +1. As previously explained, at SVM 1, both classes 2 (G) and 3(N) are considered a single class, labelled -1.

6 According to your dataset in Q3 and the design of your multi-class classifier in Q5, identify the support vectors of the linear SVMs by "inspection" and design their hyperplanes by hand. Show the calculations and explain the details of your design.

6.1 Support vectors by inspection

By inspection, class 1 (Y) can be separated from classes 2(G) and 3(N) by constructing a hyperplane (blue in figure 5) in the region of:

$$-0.71508652 \leq x_2 \leq 1.43168155$$

Hence, the support vectors of SVM 1 are:

$$\boldsymbol{X}_6 = \begin{bmatrix} -3.22162231 \\ 1.43168155 \end{bmatrix}, \boldsymbol{X}_{18} = \begin{bmatrix} -2.02275279 \\ -0.71508652 \end{bmatrix}$$

By inspection, class 2 (G) can be separated from class 3 (N) by constructing a hyperplane (red in figure 5) in the region of:

$$-0.76385808 \leq x_1 \leq 0.71393097$$

Hence, the support vectors of SVM 2 are:

$$\boldsymbol{X}_{13} = \begin{bmatrix} -0.76385808 \\ -0.85498959 \end{bmatrix}, \boldsymbol{X}_{26} = \begin{bmatrix} 0.71393097 \\ -0.83816387 \end{bmatrix}$$

The below figure helps visualize this:

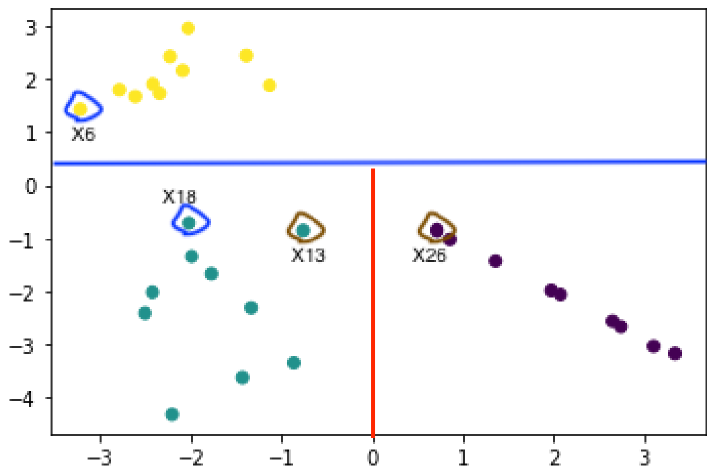

Figure 5: Support vectors

6.2 Hyperplanes by hand

6.2.1 SVM 1

As previously discussed, SVM1 separates class 1 (Y) from classes 2 (G), 3 (N). We will consider the combination of classes 2 (G), 3 (N) to be a class 4 for simplicity and reference.
We define the label for class 1 (Y) as +1, and for class 4 as -1.

$$y_6 = 1, y_{18} = -1$$

$$\frac{\partial \mathscr{L}(\mathbf{w}, w_0, \lambda)}{\partial \mathbf{w}} = \mathbf{0} \quad \Rightarrow \quad \mathbf{w} = \sum_{i=1}^{N} \lambda_i y_i \mathbf{x}_i \tag{1}$$

$$\frac{\partial \mathscr{L}(\mathbf{w}, w_0, \lambda)}{\partial w_0} = 0 \quad \Rightarrow \quad \sum_{i=1}^{N} \lambda_i y_i = 0 \tag{2}$$

Figure 6: Equations

Hyperplane:

$$\boldsymbol{w}^t \boldsymbol{x} + w_0 = 0$$

Using (1) from figure 6,

$$\boldsymbol{w} = \lambda_6 y_6 \boldsymbol{x}_6 + \lambda_{18} y_{18} \boldsymbol{x}_{18}$$

$$\boldsymbol{w} = \lambda_6 \begin{bmatrix} -3.22162231 \\ 1.43168155 \end{bmatrix} - \lambda_{18} \begin{bmatrix} -2.02275279 \\ -0.71508652 \end{bmatrix}$$

Considering that

$$y_i(\boldsymbol{w}^t \boldsymbol{x} + w_0) = 1$$

when $\boldsymbol{x}$ is a support vector, for:

$$x = x_6, y_6 = +1$$
$$y_6(\boldsymbol{w}^t \boldsymbol{x} + w_0) = 1$$

$$1 \times \left[\lambda_6 \begin{bmatrix} -3.22162231 \\ 1.43168155 \end{bmatrix} - \lambda_{18} \begin{bmatrix} -2.02275279 \\ -0.71508652 \end{bmatrix} \right]^t \begin{bmatrix} -3.22162231 \\ 1.43168155 \end{bmatrix} + w_0 = 1$$

$$\begin{bmatrix} -3.22162231\lambda_6 + 2.02275279\lambda_{18} \\ 1.43168155\lambda_6 + 0.71508652\lambda_{18} \end{bmatrix}^t \begin{bmatrix} -3.22162231 \\ 1.43168155 \end{bmatrix} + w_0 = 1$$

$$12.428562395485452\lambda_6 - 5.492769343845637\lambda_{18} + w_0 = 1$$

For:

$$x = x_{18}, y_{18} = -1$$
$$y_{18}(\boldsymbol{w}^t \boldsymbol{x} + w_0) = 1$$

$$-1 \times \left[\lambda_6 \begin{bmatrix}-3.22162231\\1.43168155\end{bmatrix} - \lambda_{18}\begin{bmatrix}-2.02275279\\-0.71508652\end{bmatrix}\right]^t \begin{bmatrix}-2.02275279\\-0.71508652\end{bmatrix} + w_0 = 1$$

$$\begin{bmatrix}-3.22162231\lambda_6 + 2.02275279\lambda_{18}\\1.43168155\lambda_6 + 0.71508652\lambda_{18}\end{bmatrix}^t \begin{bmatrix}-2.02275279\\-0.71508652\end{bmatrix} + w_0 = -1$$

$$5.492769343845637\lambda_6 - 4.602877587701994\lambda_{18} + w_0 = -1$$

Using (2) from figure 6,

$$\lambda_6 - \lambda_{18} = 0$$

$$\begin{bmatrix}12.428562395485452 & -5.492769343845637 & 1\\5.492769343845637 & -4.602877587701994 & 1\\1 & -1 & 0\end{bmatrix}\begin{bmatrix}\lambda_6\\\lambda_{18}\\w_0\end{bmatrix} = \begin{bmatrix}1\\-1\\0\end{bmatrix}$$

$$\lambda_6 = \lambda_{18} = 0.33080262$$

$$w_0 = -1.29437853$$

Using (1) from figure 6,

$$\boldsymbol{w} = 0.33080262\begin{bmatrix}-3.22162231\\1.43168155\end{bmatrix} - 0.33080262\begin{bmatrix}-2.02275279\\-0.71508652\end{bmatrix} = \begin{bmatrix}-0.39658918\\0.71015651\end{bmatrix}$$

$$w_1 = -0.39658918$$

$$w_2 = 0.71015651$$

The hyperplane of SVM1 is:

$$\boldsymbol{w}^t\boldsymbol{x} + w_0 = 0$$

$$\begin{bmatrix}-0.39658918 & 0.71015651\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} - 1.29437853 = 0$$

6.2.2 SVM 2

As previously discussed, SVM2 separates class 2 (G) from class 3 (N).
We define the label for class 2 (G) as -1, and for class 3 (N) as +1.

$$y_{26} = 1, y_{13} = -1$$

Hyperplane:

$$\boldsymbol{w}^t\boldsymbol{x} + w_0 = 0$$

Using (1) from figure 6,

$$\boldsymbol{w} = \lambda_{26}y_{26}\boldsymbol{x}_{26} + \lambda_{13}y_{13}\boldsymbol{x}_{13}$$

$$\boldsymbol{w} = \lambda_{26}\begin{bmatrix}0.71393097\\-0.83816387\end{bmatrix} - \lambda_{13}\begin{bmatrix}-0.76385808\\-0.85498959\end{bmatrix}$$

Considering that

$$y_i(\boldsymbol{w}^t\boldsymbol{x} + w_0) = 1$$

when $\boldsymbol{x}$ is a support vector, for:

$$x = x_{26}, y_{26} = +1$$

$$y_{26}(\boldsymbol{w}^t\boldsymbol{x} + w_0) = 1$$

$$1 \times \left[\lambda_{26}\begin{bmatrix}0.71393097\\-0.83816387\end{bmatrix} - \lambda_{13}\begin{bmatrix}-0.76385808\\-0.85498959\end{bmatrix}\right]^t \begin{bmatrix}0.71393097\\-0.83816387\end{bmatrix} + w_0 = 1$$

$$\begin{bmatrix}0.71393097\lambda_{26} + 0.76385808\lambda_{13}\\-0.83816387\lambda_{26} + 0.85498959\lambda_{13}\end{bmatrix}^t \begin{bmatrix}0.71393097\\-0.83816387\end{bmatrix} + w_0 = 1$$

$$1.2122160976965928\lambda_{26} - 0.17127943586084815\lambda_{13} + w_0 = 1$$

For:

$$x = x_{13}, y_{13} = -1$$

$$y_{13}(\boldsymbol{w}^t\boldsymbol{x} + w_0) = 1$$

$$-1 \times \left[\lambda_{26}\begin{bmatrix}0.71393097\\-0.83816387\end{bmatrix} - \lambda_{13}\begin{bmatrix}-0.76385808\\-0.85498959\end{bmatrix}\right]^t \begin{bmatrix}-0.76385808\\-0.85498959\end{bmatrix} + w_0 = 1$$

$$\begin{bmatrix}0.71393097\lambda_{26} + 0.76385808\lambda_{13}\\-0.83816387\lambda_{26} + 0.85498959\lambda_{13}\end{bmatrix}^t \begin{bmatrix}-0.76385808\\-0.85498959\end{bmatrix} + w_0 = -1$$

$$0.17127943586084815\lambda_{26} - 1.314486370240288\lambda_{13} + w_0 = -1$$

Using (2) from figure 6,

$$\lambda_{26} - \lambda_{13} = 0$$

$$\begin{bmatrix}1.2122160976965928 & -0.17127943586084815 & 1\\0.17127943586084815 & -1.314486370240288 & 1\\1 & -1 & 0\end{bmatrix}\begin{bmatrix}\lambda_{26}\\\lambda_{13}\\w_0\end{bmatrix} = \begin{bmatrix}1\\-1\\0\end{bmatrix}$$

$$\lambda_{26} = \lambda_{13} = 0.91569071$$

$$w_0 = 0.04682397$$

Using (1) from figure 6,

$$\boldsymbol{w} = 0.91569071\begin{bmatrix}0.71393097\\-0.83816387\end{bmatrix} - 0.91569071\begin{bmatrix}-0.76385808\\-0.85498959\end{bmatrix} = \begin{bmatrix}1.35319771\\0.01540716\end{bmatrix}$$

$$w_1 = 1.35319771$$

$$w_2 = 0.01540716$$

The hyperplane of SVM2 is:

$$\boldsymbol{w}^t\boldsymbol{x} + w_0 = 0$$

$$\begin{bmatrix}1.35319771 & 0.01540716\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} + 0.04682397 = 0$$

7 Produce a test dataset by averaging the samples for each row in Table 2, i.e., (sam- ple of class 1 + sample of class 2 + sample of class 3)/3. Summarise the results in the form of Table 3, where N is the number of SVMs in your design and "Classi- fication" is the class determined by your multi-class classifier. Explain how to get the "Classification" column using one test sample. Show the calculations for one or two samples to demonstrate how to get the contents in the table.

[-0.80322302 -0.10991796] is in Class 2 (G)
[-0.83552826 -0.12563344] is in Class 2 (G)
[-0.36885394 -0.33698234] is in Class 2 (G)
[0.16448753 -1.19307089] is in Class 3 (N)
[-0.29224086 -1.66925011] is in Class 2 (G)
[-1.64446856 -0.47555155] is in Class 2 (G)
[-1.15346971 -1.36084375] is in Class 2 (G)
[-0.3363224 -0.53050322] is in Class 2 (G)
[-0.69494663 -0.85202267] is in Class 2 (G)
[-0.71431317 -0.64212292] is in Class 2 (G)

For [-0.80322302 -0.10991796]:
Input into SMV 1 =

$$\boldsymbol{w}^t\boldsymbol{x} + w_0 = 0$$

$$\begin{bmatrix}-0.39658918 & 0.71015651\end{bmatrix}\begin{bmatrix}-0.80322302\\-0.10991796\end{bmatrix} - 1.29437853 = -1.1$$

So, we input into SVM 2 =

$$\boldsymbol{w}^t\boldsymbol{x} + w_0 = 0$$

$$\begin{bmatrix}1.35319771 & 0.01540716\end{bmatrix}\begin{bmatrix}-0.80322302\\-0.10991796\end{bmatrix} + 0.04682397 = -1.04$$

Hence, [-0.80322302 -0.10991796] is classified as Grey in Class 2.