Course Work 3

Johannes Janousek
1878155

27th of March, 2019

Contents

1 Q1

$S_1, S_2, S_3, S_4, S_5, S_6, S_7 = 1878155$

2 Q2

R_1	Method
0	One against one
1	One against all
2	Binary decision tree
3	Binary coded

The remainder of my student id was 3 and hence assigned me to the binary coded SVM.

3 Q3

Class 1	Class 2	Class 3
[1, 2]	[1, 9]	[8, 7]
[1, 1]	[2, 10]	[8, 5]
[1, 3]	[2, 8]	[9, 6]
[2, 3]	[2, 9]	[9, 7]
[2, 1]	[3, 9]	[9, 5]
[2, 2]	[1, 10]	[10, 5]
[3, 1]	[3, 11]	[10, 7]
[3, 3]	[3, 10]	[8, 6]
[4, 2]	[4, 9]	[10, 6]
[3, 2]	[2, 12]	[7, 6]
[2, 4]	[2, 11]	[11, 6]

4 Q4

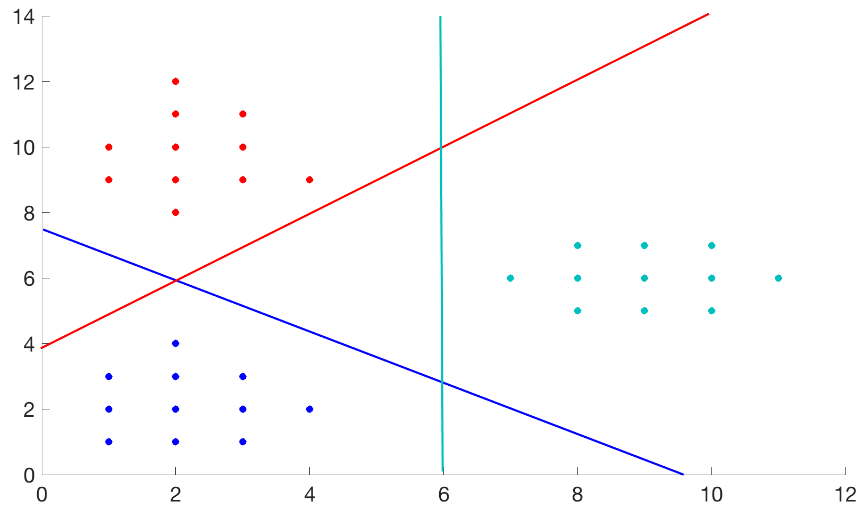

Figure 1: Three linearly separable classes.

Linear separability between sets of points is a property that occurs when at least one hyperplane exists that completely separates one set of points from another set of points. When multiple hyperplanes exist, a reasonable argument for the optimum hyperplane would be to select the one that creates the largest separation or margin between two sets. Essentially, a well placed hyperplane maximizes the distance from it to the nearest data point from each set. In the case of the generated dataset with three classes in Q3, the dataset is linearly separable if there exists at least one hyperplane that separates a given set from all the points of the two remaining sets. From the figure visualizing the generated dataset, it can be established that the dataset possesses the property of linear separability as the convex hulls of the classes do not overlap when taken in an one vs all approach.

5 Q5

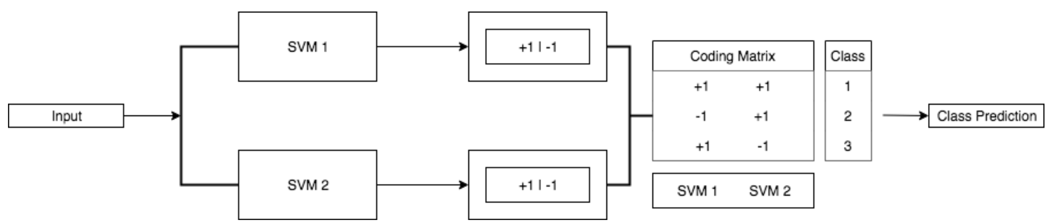

Figure 2: Block diagram of binary coded SVM.

A binary coded multi-class support vector machine approach requires $\lceil \log_2 R \rceil$ SVMs where R is the number of classes. For $\lceil \log_2 3 \rceil$ the architecture requires 2 SVMs. As can be seen in the figure, the model takes a single observation as input. SVM 1 and SVM 2 then both classify the input simultaneously. Next, in accordance to the binary coding as can be seen in the coding matrix, the class corresponding to the outputs from SVM 1 and SVM 2 is returned as the prediction output. An advantage of this multi-class classifier architecture is that the number of SVMs is reduced which in effect simplifies the evaluation of the decision making process. However, this architecture may increase overall training time as the whole dataset is used for training each individual SVM. Further, when adjusting the number of SVMs, all SVMs have to be retrained. Another drawback associated with the binary coded architecture is the possibility for predicting undefined classes when the number for binary coded strings exceed the number of classifiable classes ($2^{Number\ of\ classes}$).

6 Q6

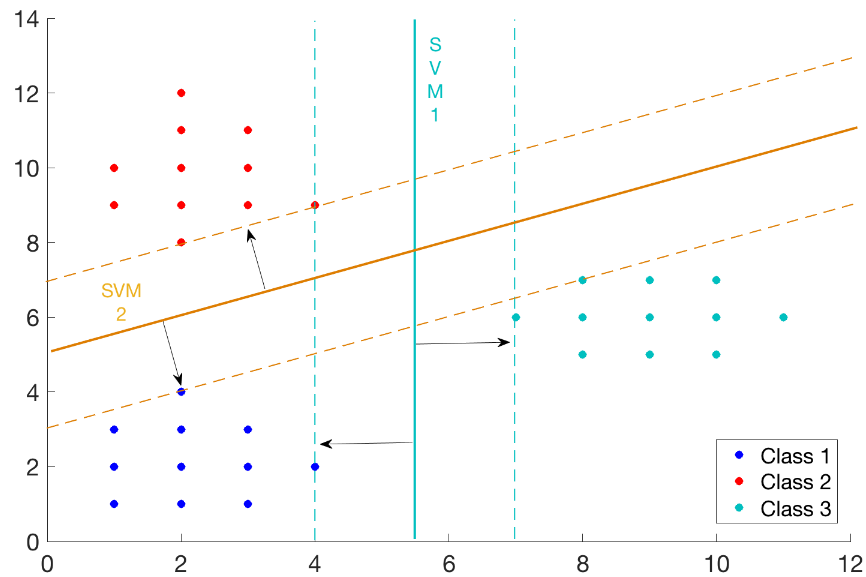

Figure 3: Support vectors and hyperplanes of the linear SVMs.

By inspection, class 3 can be separated by constructing a hyperplane in the region of $4 \leq x_1 \leq 7$. The support vectors for SVM1 are:

$$x_1 = \begin{bmatrix} 7 \\ 6 \end{bmatrix} \quad x_2 = \begin{bmatrix} 4 \\ 9 \end{bmatrix} \quad x_3 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

The labels for class 1 and class 2 are:

$$y_1 = -1 \quad y_2 = y_3 = +1$$

And since the hyperplane is $w^T x + w_0 = 0$:

$$w = -\lambda_1 \begin{bmatrix} 7 \\ 6 \end{bmatrix} + \lambda_2 \begin{bmatrix} 4 \\ 9 \end{bmatrix} + \lambda_3 \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

Further, when x is a support vector then $y_i \left(\mathbf{w}^T \mathbf{x} + w_0 \right) = 1$

$x = x_1 \quad y_1 = -1$:

$$y_1\left(\mathbf{w}^T\mathbf{x}_1 + w_0\right) = -1\times\left(\left(-\lambda_1\begin{bmatrix}7\\6\end{bmatrix} + \lambda_2\begin{bmatrix}4\\9\end{bmatrix} + \lambda_3\begin{bmatrix}4\\2\end{bmatrix}\right)^T\begin{bmatrix}7\\6\end{bmatrix} + w_0\right) = 1$$

$\Rightarrow -85\lambda_1 + 82\lambda_2 + 40\lambda_3 + w_0 = -1$

$x = x_2 \quad y_2 = 1$:

$$y_2\left(\mathbf{w}^T\mathbf{x}_1 + w_0\right) = 1\times\left(\left(-\lambda_1\begin{bmatrix}7\\6\end{bmatrix} + \lambda_2\begin{bmatrix}4\\9\end{bmatrix} + \lambda_3\begin{bmatrix}4\\2\end{bmatrix}\right)^T\begin{bmatrix}4\\9\end{bmatrix} + w_0\right) = 1$$

$\Rightarrow -82\lambda_1 + 97\lambda_2 + 34\lambda_3 + w_0 = 1$

$x = x_3 \quad y_3 = 1$:

$$y_3\left(\mathbf{w}^T\mathbf{x}_1 + w_0\right) = 1\times\left(\left(-\lambda_1\begin{bmatrix}7\\6\end{bmatrix} + \lambda_2\begin{bmatrix}4\\9\end{bmatrix} + \lambda_3\begin{bmatrix}4\\2\end{bmatrix}\right)^T\begin{bmatrix}4\\2\end{bmatrix} + w_0\right) = 1$$

$\Rightarrow -40\lambda_1 + 34\lambda_2 + 20\lambda_3 + w_0 = 1$

$$\begin{bmatrix}-85 & 82 & 40 & 1\\-82 & 97 & 34 & 1\\-40 & 34 & 20 & 1\\-1 & 1 & 1 & 0\end{bmatrix}\begin{bmatrix}\lambda_1\\\lambda_2\\\lambda_3\\w_0\end{bmatrix} = \begin{bmatrix}-1\\1\\1\\0\end{bmatrix}$$

$\lambda_1 = 0.2222 \quad \lambda_2 = 0.127 \quad \lambda_3 = 0.0952 \quad w_0 = 3.6667$

$$\mathbf{w} = \lambda_1 y_1\mathbf{x}_1 + \lambda_2 y_2\mathbf{x}_2 + \lambda_3 y_3\mathbf{x}_3 \Rightarrow \mathbf{w} = \begin{bmatrix}-0.6667\\0.0952\end{bmatrix}$$

Hyperplane: $\mathbf{w}^T\mathbf{x} + w_0 =$

$$\begin{bmatrix}-0.6667 & 0.0952\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} + 3.6667 = -0.6667x_1 + 0.0952x_2 + 3.6667 = 0$$

Margin of SVM1:

$$\frac{2}{\|\mathbf{w}\|} = \frac{2}{\sqrt{w_1^2 + w_2^2}} = \frac{2}{\sqrt{(-0.6667)^2 + 0.0952^2}} = 2.9698$$

By further inspection, a second hyperplane can be constructed by using the support vectors of class 2 on the one side and the support vectors of class 1 and class 3 on the other side. The support vectors for SVM2 are:

$$x_1 = \begin{bmatrix} 2 \\ 8 \end{bmatrix} \quad x_2 = \begin{bmatrix} 4 \\ 9 \end{bmatrix} \quad x_3 = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \quad x_4 = \begin{bmatrix} 8 \\ 7 \end{bmatrix}$$

The labels for class 1 and class 2 are:

$$y_1 = y_2 = -1 \quad y_3 = y_4 = +1$$

$$w = -\lambda_1 \begin{bmatrix} 2 \\ 8 \end{bmatrix} - \lambda_2 \begin{bmatrix} 4 \\ 9 \end{bmatrix} + \lambda_3 \begin{bmatrix} 2 \\ 4 \end{bmatrix} + \lambda_4 \begin{bmatrix} 8 \\ 7 \end{bmatrix}$$

As the steps are the same as in SVM1, I will skip to the resulting matrix representation:

$$\begin{bmatrix} -68 & -89 & 36 & 72 & 1 \\ -80 & -97 & 44 & 95 & 1 \\ -36 & -44 & 20 & 44 & 1 \\ -72 & -95 & 44 & 113 & 1 \\ -1 & -1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ w_0 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

$$\mathbf{w} = \lambda_1 y_1 \mathbf{x}_1 + \lambda_2 y_2 \mathbf{x}_2 + \lambda_3 y_3 \mathbf{x}_3 + \lambda_4 y_4 \mathbf{x}_4 \Rightarrow \mathbf{w} = \begin{bmatrix} 0.25 \\ -0.5 \end{bmatrix}$$

Hyperplane for SVM2: $\mathbf{w}^T \mathbf{x} + w_0 =$

$$\begin{bmatrix} 0.25 & -0.5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 2.5 = 0.25x_1 - 0.5x_2 + 2.5 = 0$$

Margin of SVM2:

$$\frac{2}{\|\mathbf{w}\|} = \frac{2}{\sqrt{w_1^2 + w_2^2}} = \frac{2}{\sqrt{0.25^2 + (-0.5)^2}} = 3.5777$$

7 Q7

Test Sample	Output of SVM1	Output of SVM2	Classification
[3.3333, 6]	1	1	Class 1
[3.6667, 5.3333]	1	1	Class 1
[4, 5.6667]	1	1	Class 1
[4.3333, 6.3333]	1	1	Class 1
[4.6667, 5.0000]	1	1	Class 1
[4.3333, 5.6667]	1	1	Class 1
[5.3333, 6.3333]	1	1	Class 1
[4.6667, 6.3333]	1	1	Class 1
[6, 5.6667]	1	1	Class 1
[4, 6.6667]	1	1	Class 1
[5, 7]	1	1	Class 1

The results in the Classification column can be obtained by using a hard classifier for each SVM:

$$f(\mathbf{x}) = \operatorname{sgn}\left(\mathbf{w}^T\mathbf{x} + w_0\right) \text{ where } \operatorname{sgn}(z) = \begin{cases} -1 & \text{if } z < 0 \\ +1 & \text{if } z \geq 0 \end{cases}$$

And a binary coded matrix:

SVM1	SVM2	Class
1	1	1
1	-1	2
-1	1	3

SVM1:

$$x_1 = \begin{bmatrix} 3.3333 \\ 6 \end{bmatrix} \quad x_2 = \begin{bmatrix} 3.6667 \\ 5.3333 \end{bmatrix} \quad w = \begin{bmatrix} -0.6667 \\ 0.0952 \end{bmatrix} \quad w_0 = 3.6667$$

Sample 1:

$$f(\mathbf{x}) = \operatorname{sgn}\left(\begin{bmatrix} -0.6667 \\ 0.0952 \end{bmatrix}^T \begin{bmatrix} 3.3333 \\ 6 \end{bmatrix} + 3.6667\right) = \operatorname{sgn}(2.0159) = 1$$

Sample 2:

$$f(\mathbf{x}) = \text{sgn}\left(\begin{bmatrix} -0.6667 \\ 0.0952 \end{bmatrix}^T \begin{bmatrix} 3.6667 \\ 5.3333 \end{bmatrix} + 3.6667\right) = \text{sgn}\,(1.7302) = 1$$

SVM2:

$$x_1 = \begin{bmatrix} 3.3333 \\ 6 \end{bmatrix} \quad x_2 = \begin{bmatrix} 3.6667 \\ 5.3333 \end{bmatrix} \quad w = \begin{bmatrix} 0.25 \\ -0.5 \end{bmatrix} \quad w_0 = 2.5$$

Sample 1:

$$f(\mathbf{x}) = \text{sgn}\left(\begin{bmatrix} 0.25 \\ -0.5 \end{bmatrix}^T \begin{bmatrix} 3.3333 \\ 6 \end{bmatrix} + 2.5\right) = \text{sgn}\,(0.3333) = 1$$

Sample 2:

$$f(\mathbf{x}) = \text{sgn}\left(\begin{bmatrix} 0.25 \\ -0.5 \end{bmatrix}^T \begin{bmatrix} 3.6667 \\ 5.3333 \end{bmatrix} + 2.5\right) = \text{sgn}\,(0.7500) = 1$$

As can be read in the binary encoded matrix, when both SVMs return a +1 as output, the sample is classified as $Class1$. Should the first SVM return a +1 and the second SVM return a -1, the sample is classified as $Class2$ and when the first SVM returns a -1 and the second a +1, the sample belongs to $Class3$. As the averaged samples fall within the bottom left corner, they are all classified as $Class1$ with SVM1=+1 and SVM2=+1.