Pattern Recognition Assignment 3

Your Name, Your ID

Q1: Write down your 7-digit student ID denoted as s1s2s3s4s5s6s7.

s1 = ;
s2 = ;
s3 = ;
s4 = ;
s5 = ;
s6 = ;
s7 = ;

Q2: Find R1 which is the remainder of $\frac{s1+s2+s3+s4+s5+s6+s7}{4}$.

$$\frac{s1+s2+s3+s4+s5+s6+s7}{4} = \frac{\quad}{4} = \frac{\quad}{4}$$

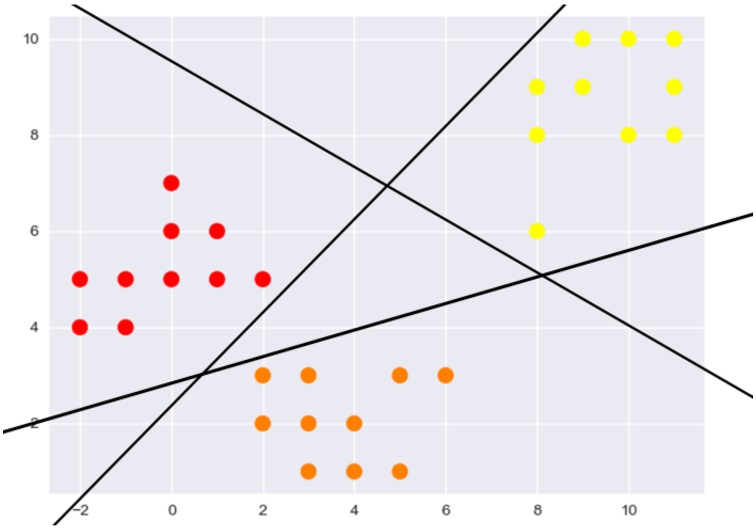
My remainder of the formula is 3, the corresponding multi-class method is **`Binary coded`**.

Q3: 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.

Data of Class 1	**Data of Class 2**	**Data of Class 3**
$(-2,\ 5)^\top$	$(2,\ 3)^\top$	$(7,\ 7)^\top$
$(-1,\ 5)^\top$	$(2,\ 2)^\top$	$(7,\ 9)^\top$
$(0,\ 5)^\top$	$(3,\ 3)^\top$	$(8,\ 9)^\top$
$(-1,\ 4)^\top$	$(3,\ 2)^\top$	$(9,\ 9)^\top$
$(-2,\ 4)^\top$	$(4,\ 1)^\top$	$(10,\ 10)^\top$
$(1,\ 5)^\top$	$(4,\ 2)^\top$	$(10,\ 8)^\top$
$(2,\ 5)^\top$	$(5,\ 3)^\top$	$(9,\ 11)^\top$
$(1,\ 6)^\top$	$(5,\ 1)^\top$	$(11,\ 8)^\top$
$(0,\ 6)^\top$	$(3,\ 1)^\top$	$(11,\ 9)^\top$
$(0,\ 7)^\top$	$(6,\ 3)^\top$	$(11,\ 10)^\top$

Table 2: Samples of Three Classes.

Q4: Plot the dataset in Q3 to show that the samples are linearly separable. Explain why your dataset is linearly separable.



We can find that all classes in the data I created can be separated by the straight lines. This meets the assumptions of the SVM.

Q5: 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 of SVMs used, class label assignment, etc.) and describe how this multi-class classifier works.

Input: 2-D data, like $(-2, 5)^\top$.
Output: Three codes, like `(+1, -1)`, This means that SVM1 classifies it as class 1, 2, SVM2 classifies it as class 2.
number of SVMs used: 3
class label assignment: For each SVM classifier, +1, -1 are two categories of binary classification.

How this multi-class classifier works:

Since there are three categories of data in Q3, we use log_2 (3) to round up to 2 SVMs for classification. Refer to the paper Solving Multiclass Learning Problems via Error-Correcting Output Codes (https://arxiv.org/pdf/cs/9501101.pdf) method, which is the same as Lecture8.pdf shows. Due to the small number of categories, I used the Exhaustive Code method, which is:

SVM 1	SVM 2
+1	+1
+1	-1
-1	+1

$$\text{SVM 1}: \underbrace{12}_{+1} | \underbrace{3}_{-1}$$

$$\text{SVM 2}: \underbrace{13}_{+1} | \underbrace{2}_{-1}$$

The above coding meets two conditions in the paper:

- **Row separation.** Each codeword should be well-separated in Hamming distance from each of the other codewords.
- **Column separation.** Each bit-position function f_i should be uncorrelated with the functions to be learned for the other bit positions $f_j, j \neq i$. This can be achieved by insisting that the Hamming distance between column i and each of the other columns be large and that the Hamming distance between column i and the *complement* of each of the other columns also be large.

The final output category is **the one with the smallest Hamming distance**(Euclidean distance can also be used) from the different categories of training data. For example, when the output is (+1, -1). The Hamming distance between (+1, -1) and the first type of encoding (+1, +1) is 2, because they have a different encoding (https://en.wikipedia.org/wiki/Hamming_distance). Similarly, the Hamming distance from the second and third types is 0 and 2. So we classify this sample into **the second category** because its predictive coding (+1, -1) has the smallest distance from the second category.

Q6: 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.

For SVM 1:

$$\frac{\partial L}{\partial w} = 0, \quad x_1 = \begin{pmatrix} 6 \\ 3 \end{pmatrix}; \quad x_2 = \begin{pmatrix} 8 \\ 6 \end{pmatrix}$$

$$\Rightarrow w = \sum_{i=1}^{n} \lambda_i y_i x_i = \lambda_1 \begin{pmatrix} 6 \\ 3 \end{pmatrix} - \lambda_2 \begin{pmatrix} 8 \\ 6 \end{pmatrix}$$

$$\frac{\partial L}{\partial b} = 0$$

$$\Rightarrow \sum_{i=1}^{n} \lambda_i y_i = 0 \Rightarrow \lambda_1 - \lambda_2 = 0$$

$$y_1(w^T x_1 + w_0) = 1 \times \left[\left(\lambda_1 \begin{pmatrix} 6 \\ 3 \end{pmatrix} - \lambda_2 \begin{pmatrix} 8 \\ 6 \end{pmatrix} \right)^T \begin{pmatrix} 6 \\ 3 \end{pmatrix} + w_0 \right]$$

$$\Rightarrow 45\lambda_1 - 66\lambda_2 + w_0 = 1$$

$$y_2(w^T x_2 + w_0) = -\left(\lambda_1 \begin{pmatrix} 6 \\ 3 \end{pmatrix} - \lambda_2 \begin{pmatrix} 8 \\ 6 \end{pmatrix} \right)^T \begin{pmatrix} 8 \\ 6 \end{pmatrix} - w_0 = 1$$

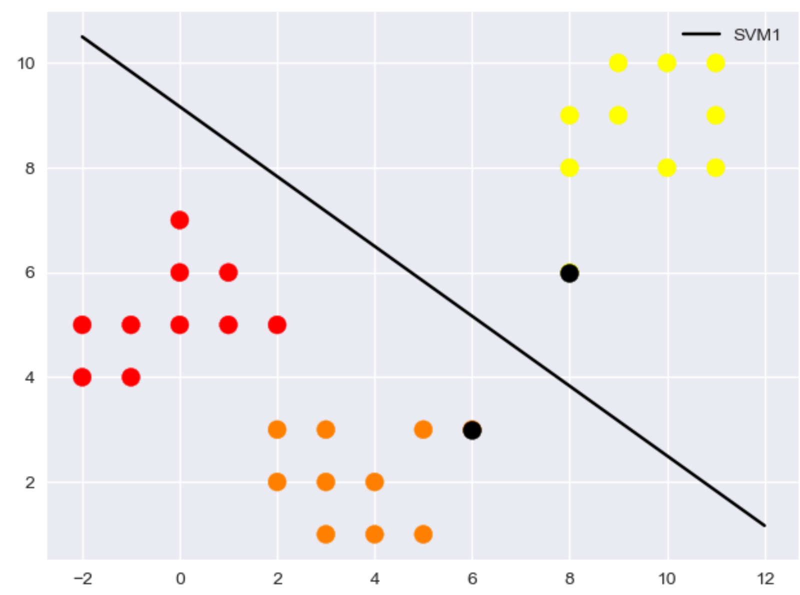
$$\Rightarrow -66\lambda_1 + 100\lambda_2 - w_0 = 1$$

$$\therefore \begin{cases} \lambda_1 = \lambda_2 = \frac{2}{13} \\ w_0 = \frac{59}{13} \end{cases}; \quad w = \lambda_1 x_1 - \lambda_2 x_2 = \begin{pmatrix} -\frac{4}{13} \\ -\frac{6}{13} \end{pmatrix}$$

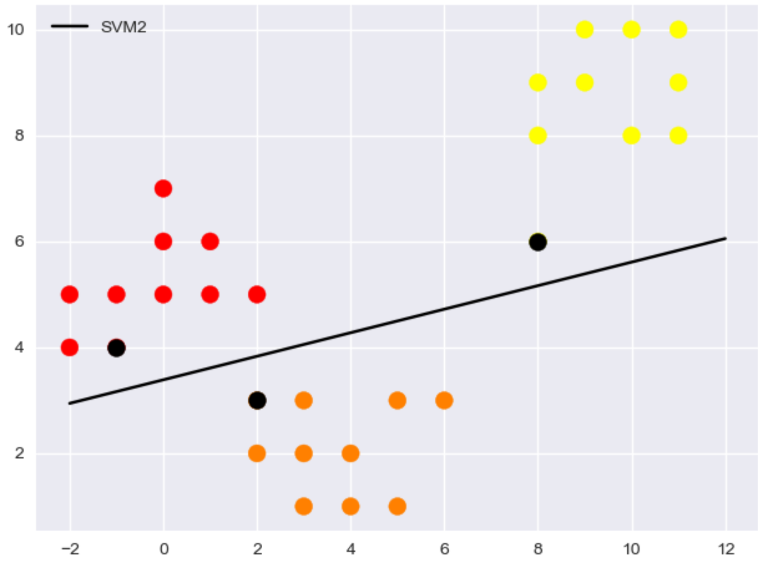
$$w^T x + w_0 = 0$$

$$\Rightarrow -\frac{4}{13} x_1 - \frac{6}{13} x_2 + \frac{59}{13} = 0$$

We can draw it with Python. The image shows the classification results of the Binary coded approach for the SVM 1, where black identifies the support vector



Similarly, we perform similar operations on **SVM2:**
Where w = (0.27, 1.19)^T, w_0 = 4.06.



Q7: Produce a test dataset by averaging the samples for each row in Table 2, i.e., (sample 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 "Classification" 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.

Test Sample: (sample of class 1 + sample of class 2 + sample of class 3)/3

Test Sample	Output of SVM 1	Output of SVM 2	Final Classification
[2.33, 5.0]	+1	+1	1
[2.67, 5.33]	+1	+1	1
[2.67, 5.67]	+1	+1	1
[3.67, 5.0]	+1	+1	1
[4.0, 5.0]	+1	+1	1
[5.0, 5.0]	+1	+1	1
[5.33, 6.33]	-1	+1	3

[5.67, 5.0]	+1	+1	1
[4.67, 5.33]	+1	+1	1
[5.67, 6.67]	-1	+1	3

Table 3: Summary of Classification Accuracy.

How to get the "Classification" column using one test sample:

We substitute each data into the hyperplane equation given in **Problem 6**.
Noticed:

- $$\begin{cases} \mathbf{w}^T\mathbf{x} + w_0 \geq 1, & \forall \mathbf{x} \in \text{class 1 (``}+1\text{'')} \\ \mathbf{w}^T\mathbf{x} + w_0 \leq -1, & \forall \mathbf{x} \in \text{class 2 (``}-1\text{'')} \end{cases}$$

For example, sample 1 is $[2.33, 5.0]$, and for SVM 1 we have:

$$(-\frac{4}{13}, -\frac{6}{13} \cdot (2.33, 5.0)^\top + \frac{59}{13} = 1.5138 \geq +1$$

So SVM 1 classified sample 1 as +1 (class 1 and class 2). The same calculations for other samples can be obtained.