HOMEWORK 4

TASK 1:

Q1) **Margin:** In a Support Vector Machine, a margin is the space between the hyperplane and the decision boundaries. The idea is to maximize the size of the margin between the data points and the hyperplane. Maximizing margin is equivalent to minimizing loss.

Q2) **A kernel function/ kernel trick** provides the solution when handling non-separable dataset. It does so by creating higher dimensions of the original dataset.

Q3) Kernel methods are a class of algorithms that are mainly used for pattern analysis. Kernels provide a shortcut for avoiding complex calculations. In essence, a kernel function is a mathematical trick that allows the SVM to perform a 'two-dimensional' classification of a set of original on-dimensional data. That is, a kernel function projects data from a low-dimensional space to a space of higher dimensions.

Q4) Feature vectors are combined with weights using a dot product in order to construct a linear predictor function that is used to determine the score for making a prediction. In a kernel trick, the ultimate benefit is that the objective function that we optimize, in order to fit into higher dimensional decision boundary only includes the dot product of the transformed feature vectors. Given a kernel function, K (x_i, x_j), the feature vectors are given by $\Phi(x_i)$, $\Phi(x_j)$.

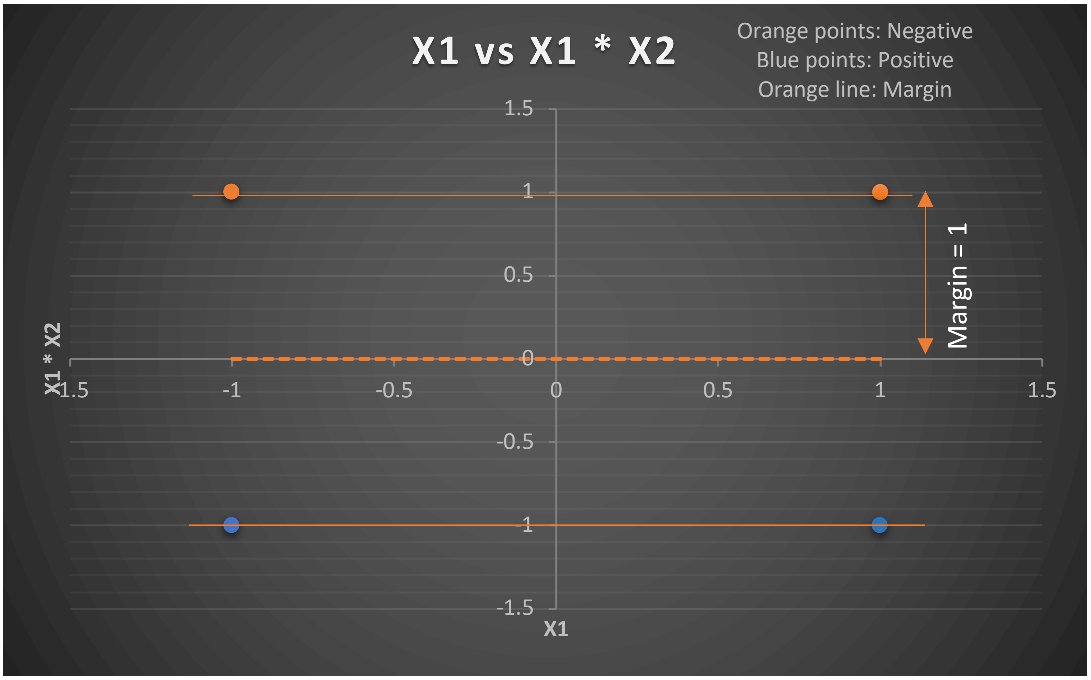
TASK 2:
Original data: [x1, x2]

X1	**X2**	**Output**
-1	-1	Negative
-1	+1	Positive
+1	-1	Positive
+1	+1	Negative

Mapped Data: [x1, x1 * x2]

X1	**X1 * X2**	**Output**
-1	+1	Negative
-1	-1	Positive
+1	-1	Positive
+1	+1	Negative

Plotting it in a graph:



Blue points are those that have positive output and Orange dots are those that have negative outputs. The orange lines represent the positive and negative margin, while the orange dotted line represents maximum margin separating hyperplane. Margin = 1 unit.

TASK 3:

Given, $(x1 - a)^2 + (x2 - b)^2 - r^2 = 0$
Expand using the formula: $(a - b)^2 = a^2 + b^2 - 2ab$
Hence, $x1^2 + a^2 - 2ax1 + x2^2 + b^2 - 2bx2 - r^2 = 0$
$x1^2 + x2^2 - 2ax1 - 2bx2 = r^2 - a^2 - b^2$

Features: $x1^2$, $x2^2$, x1, x2 and their corresponding weights are 1, 1, -2a, -2b
Let's map the variables as: $x1^2$ -> X0, $x2^2$ -> X1, x1 -> X2 and x2 -> X3

Thus, representing them as matrix multiplication, we get:
$[1, 1, -2a, -2b] * [X0, X1, X2, X3]^T$ = Constant = -B + C
Where, [1, 1, -2a, -2b] are the weights given by the vector W
And $[X0, X1, X2, X3]^T = X$

Hence, W * X = Constant = -B + X
WX + B = C
However, the SVM equation is given by Wx + b = + or – 1

Thus, we can say that every circular region is linearly separable from the rest of the plane in the feature space ($x1$, $x2$, $x1^2$, $x2^2$).

TASK 4:

Given, $c(x1 - a)^2 + d(x2 - b)^2 - 1 = 0$
Expand using the formula: $(a - b)^2 = a^2 + b^2 - 2ab$
Hence, $c(x1^2 + a^2 - 2ax1) + d(x2^2 + b^2 - 2bx2) - 1 = 0$
$cx1^2 + ca^2 - 2acx1 + dx2^2 + db^2 - 2dbx2 - 1 = 0$
$cx1^2 - 2acx1 + dx2^2 - 2dbx2 - 1 = - ca^2 - db^2$

Features: $x1^2$, $x2^2$, x1, x2, 1 and their corresponding weights are [c, d, -2ac, -2db, -1]
Let's map the variables as: 1 -> X0, x1 -> X1, x2 -> X2, $x1^2$ -> X3, $x2^2$ -> X4 and x1x2 -> X5

Thus, representing them as matrix multiplication, we get:
$[-1, -2ac, -2db, c, d, 0] * [X0, X1, X2, X3, X4, X5]^T$ = Constant
Where, [-1, -2ac, -2db, c, d, 0] are the weights given by the vector W
And $[X0, X1, X2, X3, X4, X5]^T = X$

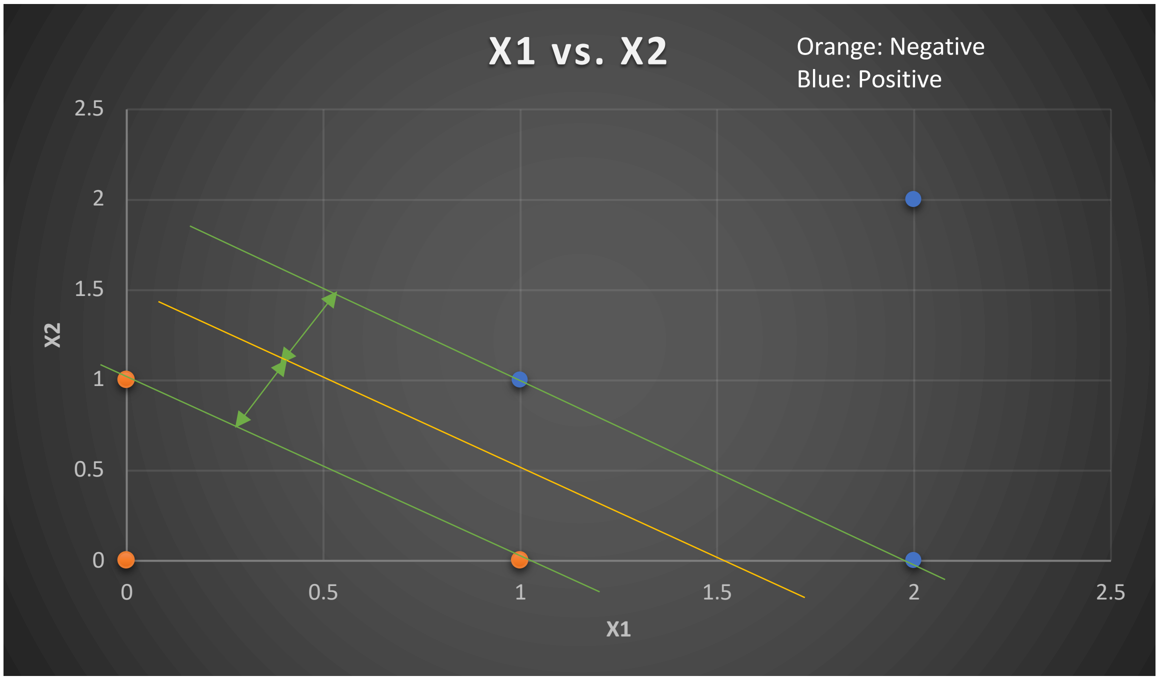
Hence, W * X = Constant = -B + X
WX + B = C
However, the SVM equation is given by Wx + b = + or – 1

If the test data point is located outside the ellipse, we give a value > 0
If the test data point is inside the ellipse, we give it a value < 0

Hence, we can say that SVMs with this kernel can separate any elliptic region from the rest of the plane.

Task 5:

Plotting the given points:



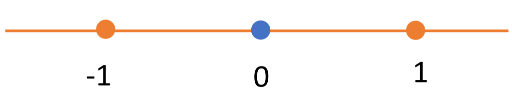
The blue dots have the positive class labels and the orange dots have the negative class labels. The green lines represent the positive and the negative hyperplane. The yellow line represents the maximum margin hyperplane (also called the maximum margin classifier).

a) The six training points are plotted, and it can be seen that the classes (+, -) are linearly separable.

b) The yellow line which is the maximum margin hyperplane is shown and the points at (x1, x2) = (1,0), (0,1), (1,1) and (2,0) are the support vectors. The maximum margin hyperplane is given by y + x = 1.5 and the weight vector is [1, 1].

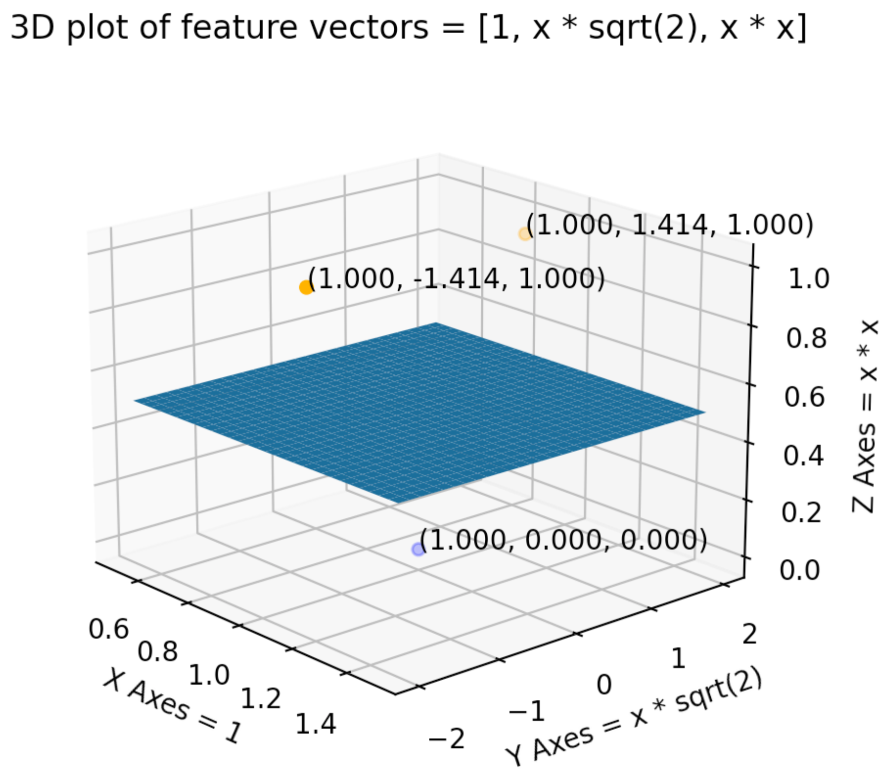
TASK 6:

a)



The Orange points have negative label and the blue point has a positive label. Based on the graph, it can be seen that the classes are not linearly separable.

b)



From the graph, the blue point represents the positive label and the orange points represent the negative label. The blue plane in the figure is the hyperplane at z = 0.5. Thus, it can be noticed that the classes are now linearly separable.

TASK 7:

The output below shows the five-fold cross validation on Titanic dataset using SVM. Linear, Quadratic and RBF kernels were used to get the below outputs

```
Average metrics over five folds for LINEAR KERNEL:

The average accuracy is: 0.7878
The average precision is: 0.7424
The average recall is: 0.6830
The average f1_score is: 0.7100

SVM LINEAR KERNEL METRICS:
The accuracy on the entire model is: 1.0000
```

```
The precision on the entire model is: 1.0000
The recall on the entire model is: 1.0000
The f1 score on the entire model is: 1.0000

Average metrics over five folds for QUADRATIC KERNEL:

The average accuracy is: 0.6588
The average precision is: 0.6917
The average recall is: 0.1830
The average f1_score is: 0.2886

SVM QUADRATIC KERNEL METRICS:

The accuracy on the entire model is: 0.6531
The precision on the entire model is: 0.5686
The recall on the entire model is: 0.1908
The f1 score on the entire model is: 0.2857

Average metrics over five folds for RBF KERNEL:

The average accuracy is: 0.7060
The average precision is: 0.6322
The average recall is: 0.5912
The average f1_score is: 0.6038

SVM RBF KERNEL METRICS:

The accuracy on the entire model is: 0.7201
The precision on the entire model is: 0.6048
The recall on the entire model is: 0.6645
The f1 score on the entire model is: 0.6332
```

Based on the output, it is found that SVM with linear kernel performs better (with 78.78% accuracy) than quadratic and RBF kernels.

GitHub link:

Assignment_04:

https://github.com/monicabernard/CAP-5610_Machine-Learning.git