CAP- 5610 Machine Learning

Homework 3

Name: Nabila Shahnaz Khan

NID: 5067496

Support Vector Machines (SVMs)

Task 1:

Q1) What is the margin and support vectors?

Answer: **Margin** is the width that the boundary could be increased by before hitting a data point. It shows the gap between the data points of two classes in an Euclidian space. **Support Vectors** are those data points that the margin pushes up against. The classification margin depends on these support vectors.

Q2) How does SVM deal with non-separable data?

Answer: SVM uses kernel trick in order to separate non-linearly separable data. It is done by applying higher dimensionality to the dataset.

Q3) What is a kernel?

Answer: Kernel function is a set of mathematical functions that take data as input and directly obtains the value of inner product.

Q4) How does a kernel relate to feature vectors?

Answer: In case of non-linearly separable feature vectors, kernel function can convert them to a new set of feature vectors that are linearly separable by introducing higher dimensionality. It can calculate the pairwise distances of the data points without actually mapping the data points to the higher dimensionality feature space. As a result, the algorithm becomes computationally less expensive.

Task 2:

Construct a support vector machine that computes the kernel function. Use four values of +1 and -1 for both inputs and outputs:

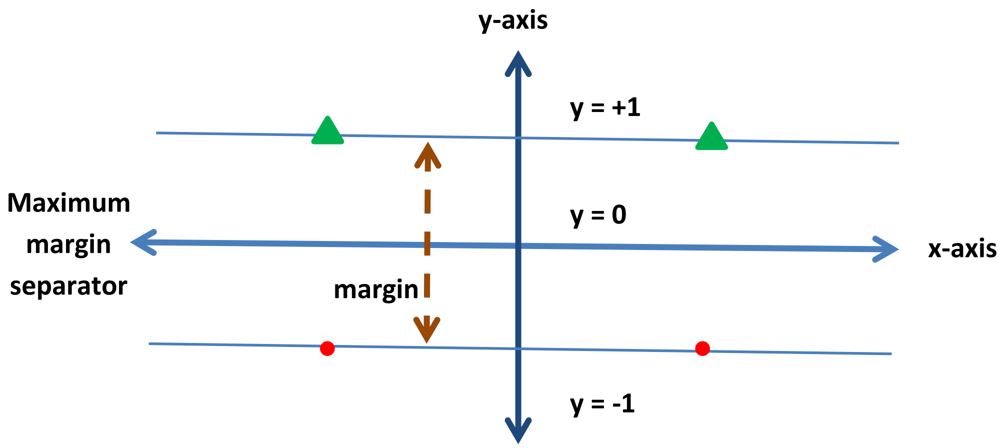
● [−1, −1] (negative)
● [−1, +1] (positive)
● [+1, −1] (positive)
● [+1, +1] (negative)

Map the input [x1, x2] into a space consisting of x1 and x1x2. Draw the four input points in this space, and the maximal margin separator. What is the margin?

Answer: After mapping the data points using kernel function, the new data points are:

Before [x1, x2]	After [x1, x1x2]
[−1, −1] (negative)	[−1, +1] (negative)
[−1, +1] (positive)	[−1, −1] (positive)
[+1, −1] (positive)	[+1, −1] (positive)
[+1, +1] (negative)	[+1, +1] (negative)

As we can see, after mapping, the set of points became linearly separable. The negative data points will always remain on the positive side of y-axis with a value of +1. And the positive data points will always remain on the negative side of y-axis with a value of -1. So, the middle separator line is y-axis itself where x1x2 = 0. And the margin width is 2.



Task 3:

Recall that the equation of the circle in the 2-dimensional plane is $(x1 - a)^2 + (x2 - b)^2 - r^2 = 0$. Please expand out the formula and show that every circular region is linearly separable from the rest of the plane in the feature space $(x1, x2, x1^2, x2^2)$.

Answer: Here, the circular equation is: $(x1 - a)^2 + (x2 - b)^2 - r^2 = 0$

After expanding:

$(x1 - a)^2 + (x2 - b)^2 - r^2 = 0$

⇨ $x1^2 - 2ax1 + a^2 + x2^2 - 2bx2 + b^2 - r^2 = 0$

⇨ $x1^2 + x2^2 - 2ax1 - 2bx2 + (a^2 + b^2 - r^2) = 0$

Here, the equation containing variables $(x1, x2, x1^2, x2^2)$ intercepts a string of constants $(a^2 + b^2 - r^2)$ which will separate the circular region from the rest of the plane. So, it can be said 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:

Recall that the equation of an ellipse in the 2-dimensional plane is $c(x1 - a)^2 + d(x2 - b)^2 - 1 = 0$. Please show that an SVM using the polynomial kernel of degree 2, $K(u, v) = (1 + u \cdot v)^2$, is equivalent to a linear SVM in the feature space ($1, x1, x2, x1^2, x2^2, x1x2$) and hence that SVMs with this kernel can separate any elliptic region from the rest of the plane.

Answer: Here, the equation of the ellipse in the 2-dimensional plane is:

$c(x1 - a)^2 + d(x2 - b)^2 - 1 = 0$

After expanding:

$c(x1 - a)^2 + d(x2 - b)^2 - 1 = 0$

⇨ $cx1^2 - 2acx1 + ca^2 + dx^2 - 2bdx2 + db^2 - 1 = 0$

⇨ $cx1^2 - 2acx1 + dx^2 - 2bdx2 + (ca^2 + db^2 - 1) = 0$

Here, the equation containing variables ($x1, x2, x1^2, x2^2$) intercepts a string of constants of the form ($a^2 + b^2 - r^2$). Ellipse equation contains the terms $x1, x2, x1^2, x2^2$ and kernel $K(u, v) = (1 + u \cdot v)^2$ also contains terms ($x1, x2, x1^2, x2^2$) in it. So, an SVM using the polynomial kernel of degree 2, $K(u, v) = (1 + u \cdot v)^2$, is equivalent to a linear SVM in the feature space ($1, x1, x2, x1^2, x2^2$, $x1x2$). Hence SVMs with this kernel can separate any elliptic region from the rest of the plane.

Task 5:

Consider the following training data:

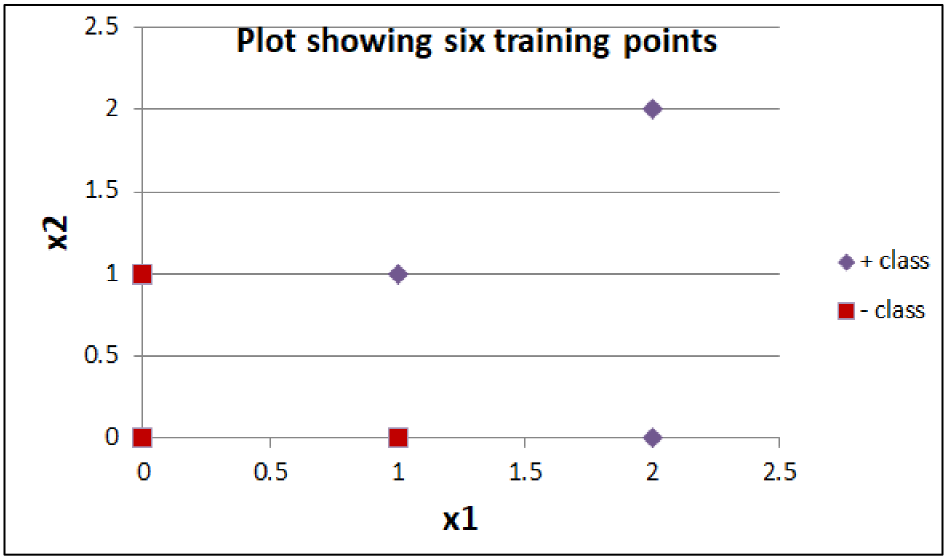
class	x_1	x_2
+	1	1
+	2	2
+	2	0
−	0	0
−	1	0
−	0	1

(a) Plot these six training points. Are the classes {+, −} linearly separable?

(b) Construct the weight vector of the maximum margin hyperplane by inspection and identify the support vectors

Answer:

a) The data points have been plotted below:



From the figure, we can clearly see that the points of the two classes are linearly separable.

b) Here, the support vectors are:

For class + : (2, 0), (1, 1)

For class - : (1, 0), (0, 1)

The center line of the margin will go through the point (1.5, 0) and it's slope will be:

Slope, m = $\frac{change\ in\ y}{change\ in\ x}$

$= \frac{1-0}{1-2}$

= - 1

So, the center line of the margin would be, x2 = -1 * x1 + 1.5; x1 + x2 = 1.5

And the weight vector is $(1, 1)^T$.

Task 6:

Consider a dataset with 3 points in 1-D:

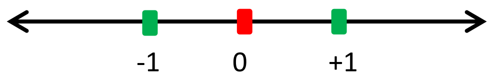
class		x
+		0
-		-1
-		+1

(a) Are the classes {+, −} linearly separable?

(b) Consider mapping each point to 3-D using new feature vectors $\phi(x)$ = [1, sqrt(2)x, x^2]. Are the classes now linearly separable? If so, find a separating hyperplane.

Answer:

a) For the given data points the plot is given below:



So, the classes are not linearly separable.

b) After mapping the data points using kernel function, the new data points are:

Before [x]	After [1, sqrt(2)x, x^2]
0 (positive)	[1, 0, 0] (positive)
-1 (negative)	[1, $-\sqrt{2}$, , 1] (negative)
+1 (negative)	[1, $\sqrt{2}$, , 1] (negative)

Suppose the three axis are X, Y, Z where the points are represented as [x, y, z]. Then the separating hyperplane could be Z = 0.5.

Task 7:

Please report your five-fold cross-validation classification accuracies on the Titanic training set, with respect to the linear, quadratic, and RBF kernels. Which kernel is the best in your case?

Answer: The accuracy, precision, recall and f1_score on the Titanic training set, with respect to the linear, quadratic, and RBF kernels using five-fold cross-validation is given below:

	Linear	Quadratic	RBF
Accuracy	79.12	61.95	68.69
Precision	74.93	60.0	67.91
Recall	68.41	0.88	34.52
f1_score	71.49	1.73	45.56

In my case, linear kernel seems to work best.

Additional Questions:

- It took me approximately 7-8 hours to complete the assignment.
- The problems were challenging because I also needed to think from mathematical aspect.
- I liked the practical implementation part most.

Code Link: https://github.com/NabilaKhan/CAP-5610-Machine-Learning-/blob/main/CAP-5610-HW3.ipynb

[Please let me know if you are having any issue to find the code]