Task 1: Q1: What is the margin and support vectors? Q2: How does SVM deal with non-separable data? Q3: What is a kernel? Q4: How does a kernel relate to feature vectors?

Ans 1)

Support vectors: These are the data points that influence the position and orientation of the Hyperplane. These are closer to the hyperplane. These points touch the boundary of the margin.

Margin: It is the distance from the decision surface to the closest data point. It is the width that the boundary could be increased by before hitting a data point.

Ans 2)

When the data is not separable, the data points are converted to higher dimensional space which makes them linearly separable. SVM uses kernel functions to transform the data points to higher dimensional space.

Ans 3)

Kernel are used by SVM algorithm which takes data as the input and transform into the required form. It transforms the training set so that non-linear decision surface can be transformed to a linear equation in higher dimensional space. Different types of kernels are Linear, Polynomial, Radial basis function (RBF) and Sigmoid.

Ans 4)

The Kernel function acts as a modified dot product for the feature vectors in the feature space. It takes input of the data and converts them to higher dimensional space.

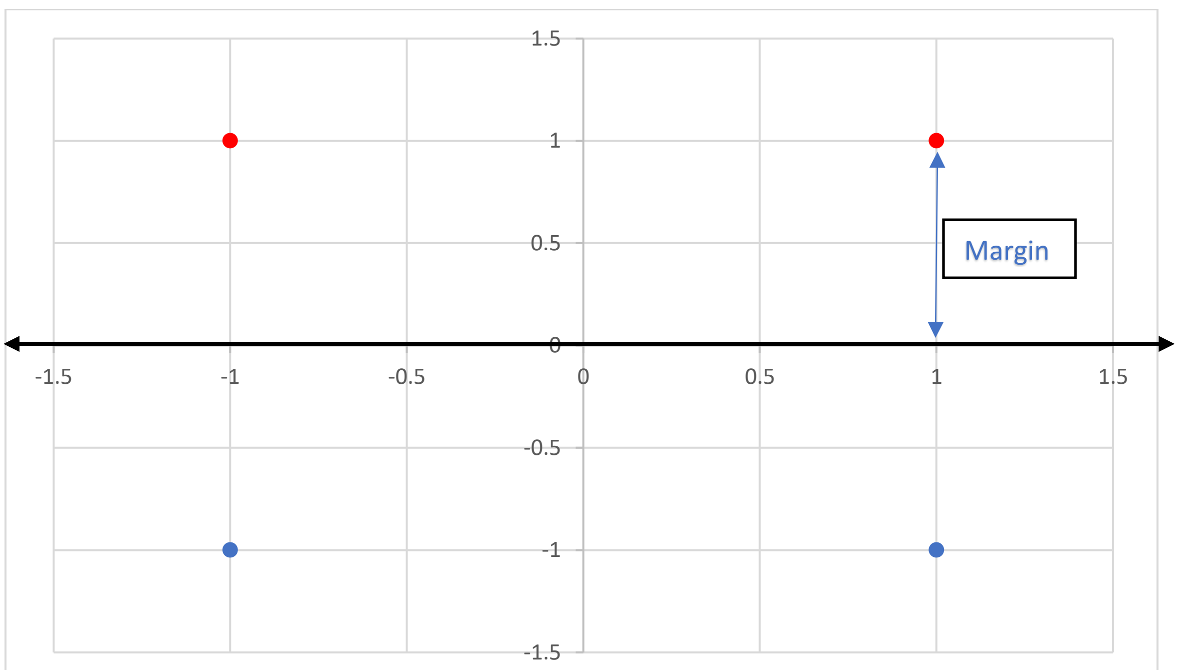
Task 2: Construct a support vector machine that computes the kernel function. Use four values of +1 and -1
for both inputs and outputs:
• [−1, −1] (negative)
• [−1, +1] (positive)
• [+1, −1] (positive)
• [+1, +1] (negative).
Map the input [x1, x2] into a space consisting of x1 and x1x2. Draw the four input points in this space, and the maximal margin separator. What is the margin?

Ans)

[x_1, x_2] (Class) => [x_1, x_1*x_2] (Class)

1) [-1, -1] (-ive) => [-1, 1] (-ive)
2) [-1, 1] (+ive) => [-1, -1] (+ive)
3) [1, -1] (+ive) => [1, -1] (+ive)
4) [1, 1] (-ive) => [1, 1] (-ive)

Mapping the inputs [x_1, x_1*x_2]:



The dark black double headed arrow line is the **Maximal margin separator**. As shown in the diagram, Margin is 1.

Task 3: Recall that the equation of the circle in the 2-dimensional plane is (x1 − a)2 + (x2 − b)2 − r2 = 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, x12, x22).

Ans)

Expanding the equation of the circle:

$(x_1 - a)^2 + (x_2 - b)^2 - r^2 = 0$
$x_1^2 + a^2 - 2ax_1 + x_2^2 + b^2 - 2bx_2 - r^2 = 0$
$x_1^2 + x_2^2 - 2ax_1 - 2bx_2 + a^2 + b^2 - r^2 = 0$
$[1, 1, -2a, -2b]. [x_1^2, x_2^2, x_1, x_2]^T + (a^2 + b^2 - r^2) = 0$

The above equation can be written into the form of the linear equation w*x + b=0, where

$w = [1, 1, -2a, -2b]$
$x = [x_1^2, x_2^2, x_1, x_2]$
$b = a^2 + b^2 - r^2$

Therefore, it can be said that circular region is linearly separable from the rest of the plane in the feature space $[x_1^2, x_2^2, x_1, x_2]$.

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 · v)2, is equivalent to a linear SVM in the feature space (1, x1, x2, x12, x22, x1x2) and hence that SVMs with this kernel can separate any elliptic region from the rest of the plane.

Ans)

i) Expanding the equation of ellipse:

$$c * (x_1 - a)^2 + d * (x_2 - b)^2 - 1 = 0$$
$$c * (x_1^2 + a^2 - 2ax_1) + d * (x_2^2 + b^2 - 2bx_2) - 1 = 0$$
$$cx_1^2 + ca^2 - 2acx_1 + dx_2^2 + db^2 - 2bdx_2 - 1 = 0$$
$$0 * (1) - 2acx_1 - 2bdx_2 + cx_1^2 + dx_2^2 + 0 * (x_1x_2) + ca^2 + db^2 - 1 = 0 \quad \text{---- \textbf{Eq (1)}}$$

ii) Working on Kernel function: $K(u, v) = (1 + u.v)^2$

Let u = x_1 = (x_{11}, x_{12})
v = x_2 = (x_{21}, x_{22})

Therefore, $\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = (1 + \mathrm{x}_1 . \mathrm{x}_2)^2$

$$= \left(1 + (x_{11}, x_{12}) . (x_{21}, x_{22})\right)^2$$
$$= (1 + x_{11}x_{21} + x_{12}x_{22})^2$$
$$= (1 + x_{11}x_{21})^2 + 2 * (1 + x_{11}x_{21}) * x_{12}x_{22} + x_{12}^2x_{22}^2$$
$$= 1 + 2x_{11}x_{21} + x_{11}^2x_{21}^2 + 2x_{12}x_{22} + 2x_{11}x_{21}x_{12}x_{22} + x_{12}^2x_{22}^2$$
$$= 1 + 2x_{11}x_{21} + 2x_{12}x_{22} + x_{11}^2x_{21}^2 + x_{12}^2x_{22}^2 + 2x_{11}x_{21}x_{12}x_{22}$$
$$= \left[1, \sqrt{2}x_{11}, \sqrt{2}x_{12}, x_{11}^2, x_{12}^2, \sqrt{2}x_{11}x_{12}\right] . \left[1, \sqrt{2}x_{21}, \sqrt{2}x_{22}, x_{21}^2, x_{22}^2, \sqrt{2}x_{21}x_{22}\right]$$

$$K(x_1, x_2) = \Phi(x_1) . \Phi(x_2)$$

Therefore, the above kernel function of degree 2 is equivalent to linear SVM in feature space $[1, x_1, x_2, x_1^2, x_2^2, x_1x_2]$ and SVM with this kernel function can separate any elliptical region.

Using Eq (1) to map on linear equation wx + b = 0:

$$w = [0, -2ac, -2bd, c, d, 0]$$
$$x = [1, x_1, x_2, x_1^2, x_2^2, x_1x_2]$$
$$b = [ca^2 + db^2 - 1]$$

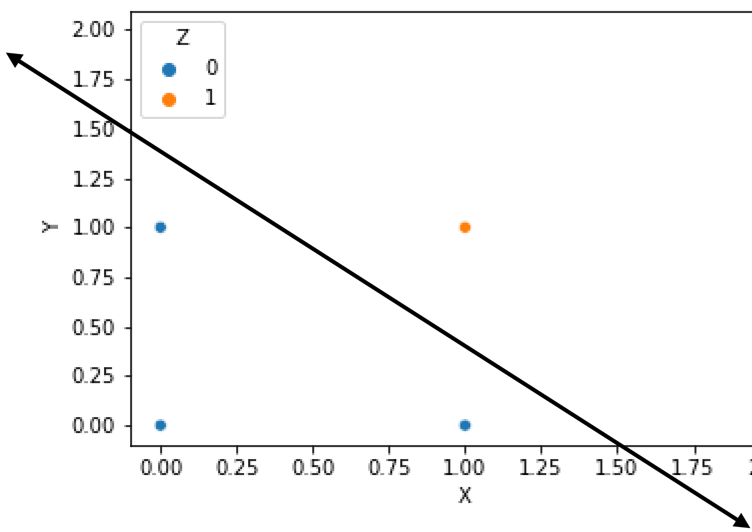
Hence the above elliptical region can be linearly separated.

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?

Ans)



As it can be seen, the above points are linearly separable.

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

Ans)

As the line passes through the points (3/2,0) & (0,3/2):

Slope of line = -1

Equation of the hyperplane: $x_1 + x_2 = 1.5$ **=>** $[1,1].[x_1, x_2] - 1.5 = 0$ --- (i)

Linear Equation: $wx + b = 0$ --- (ii)

Comparing both the equation (i) & (ii), we get:

Weight vector(w) = [1,1]

Support vectors are the data point closest to the hyperplane.

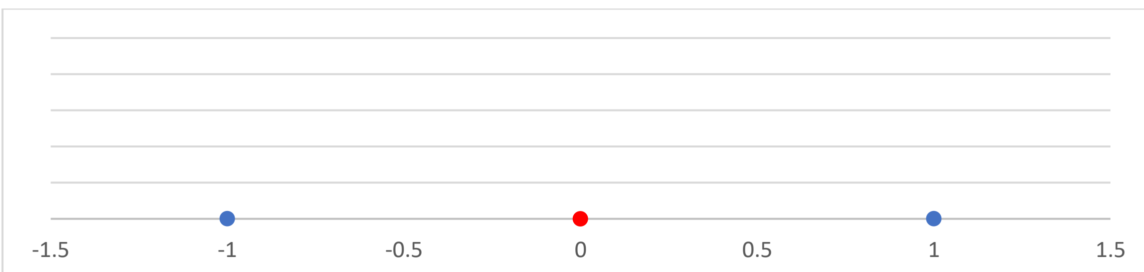
Support Vectors = [1,1], [2,0], [1,0] & [0,1] (As the perpendicular distance of these points is equal from the hyperplane).

Task 6: Consider a dataset with 3 points in 1-D:

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

(class)	x
$+$	0
$-$	-1
$-$	$+1$

Ans)



As it can be observed that the above plotted points are not linearly separable.

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

Ans)

Converting the 1-D points into the 3-D feature vectors $\Phi(x) = \left[1, \sqrt{2}x, x^2\right]$:

Class	x	1	√2 x	x^2
+	0	1	0	0
-	-1	1	-√2	1
-	1	1	√2	1

Figure 1)

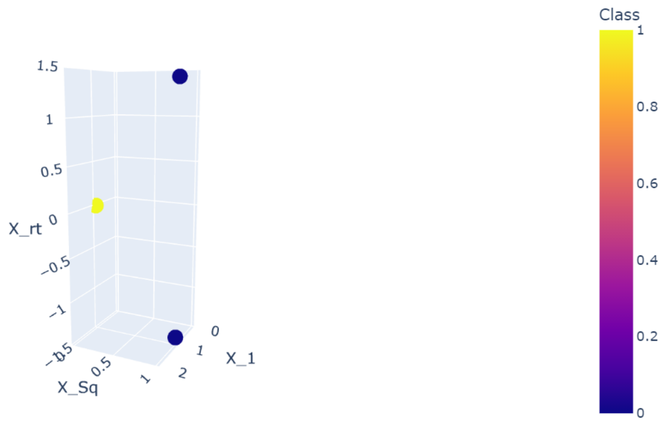
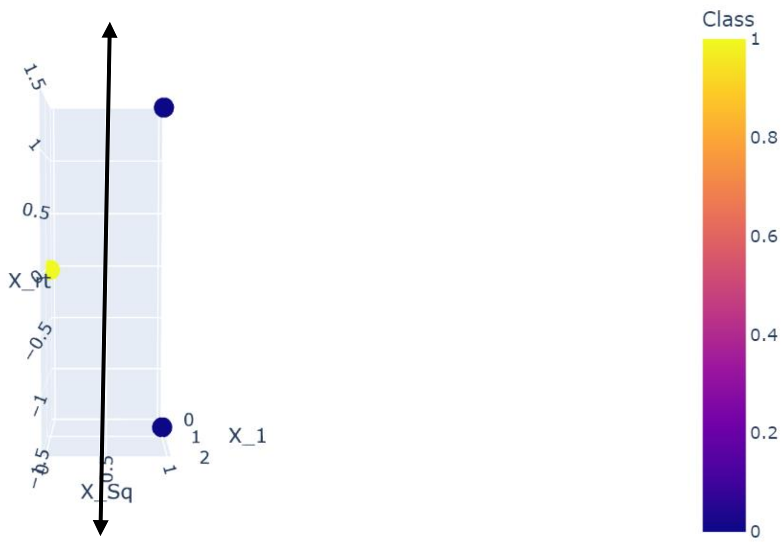


Figure 2)



Now we can see that the points are linearly separable. We can see that the hyperplane would be the one passing through :

$$x^2 = 0.5$$

Note: The program of the above plotted 3-D graph model is provided in the end of Jupyter notebook for the Task 7.

Task 7: Learning SVMs on the Titanic dataset ((https://www.kaggle.com/c/titanic). Please report your fivefold cross validation classification accuracies on Titanic training set, with respect to the linear, quadratic, and RBF kernels. Which kernel is the best in your case?

Ans)

Answer to this question is provided in the Jupyter notebook uploaded on the below GitHub link:

https://github.com/anujsarda3/ML_HW3_SVM