Task 1:

Q1: What is the margin and support vectors?

Margin refers to the distance between observations and the hyperplane

Support Vectors: The datapoints that the margin push against and touch the boundary of the margin

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

SVM projects the data that was non-separable to a high dimensional feature space where the data then becomes separable. Projection is done using kernel trick using kernels such as RBF-Kernel, Polynomial Kernel etc.

Q3: What is a Kernel?

Kernel is a function and used to define a high dimensional space.

(X) feature on Kernel function transform to (X,X^2)

Kernel function measures the correlation/distance between data points in high-dimensional space.

The data now in Higher dimensional space becomes linearly separable.

Q4: How does kernel relate to feature vectors?

Kernel function implicitly transforms feature vectors from low-dimensional space d to high dimensional space d_dash where d_dash > d.
A kernel essentially computes the dot product of two data vectors x and y in some feature space implicitly. While doing this implicitly, the kernel transforms feature vectors implicitly

For example, in polynomial kernel of degree 2,

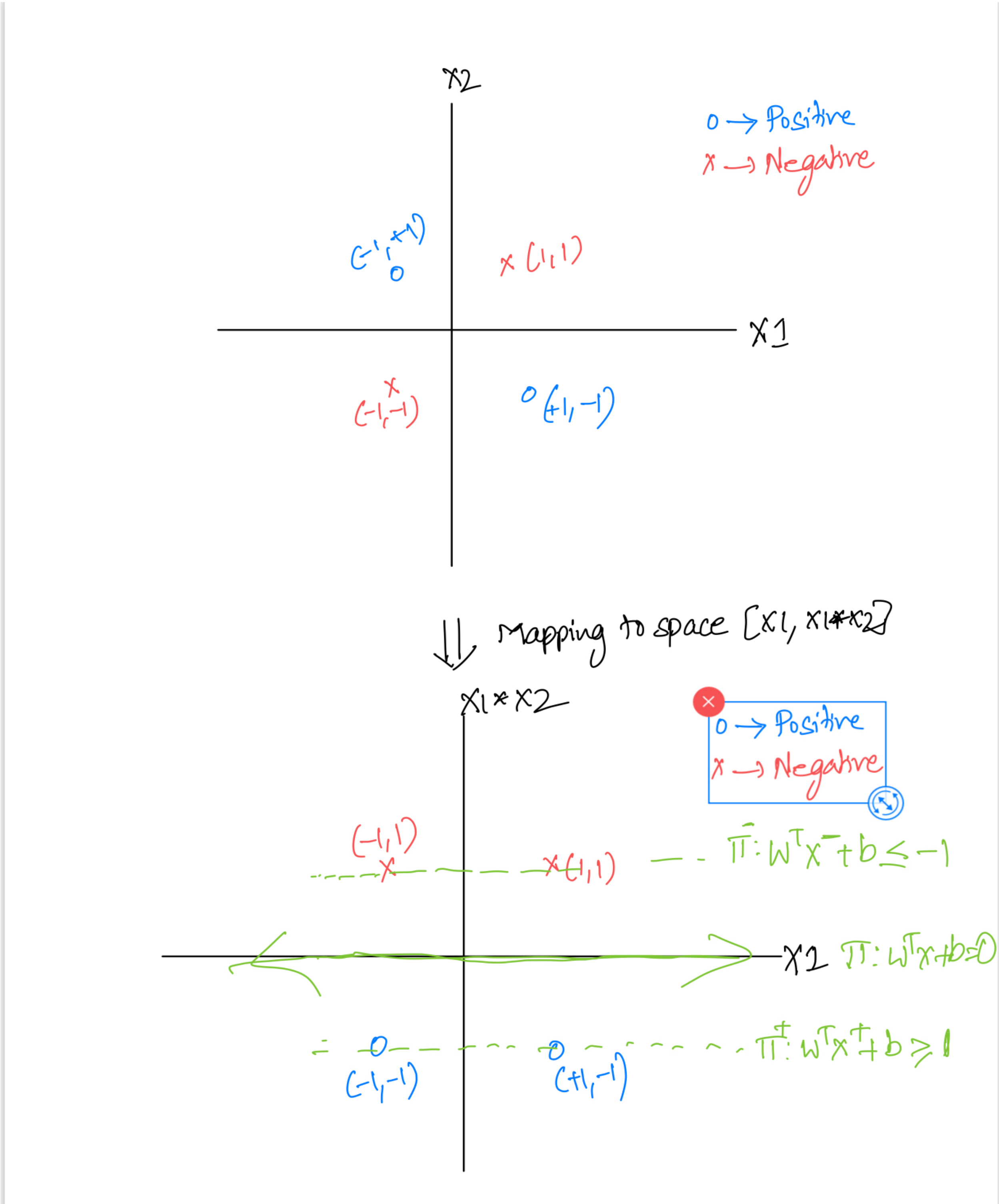
$$K(x, y) = (1 + X^T Y)\hat{}2$$

x = (x1,x2), y = (y1,y2)

$$k(\boldsymbol{x}, \boldsymbol{y}) = (1 + \boldsymbol{x}^T\boldsymbol{y})^2 = (1 + x_1y_1 + x_2y_2)^2 = 1 + x_1^2y_1^2 + x_2^2y_2^2 + 2x_1y_1 + 2x_2y_2 + 2x_1x_2y_1y_2$$

Which transforms feature vectors to new space.

TASK 2



If we use X-axis as the hyperplane, we will get low generalization error in the above case.
Margin will be 1 unit.

Task 3

The equation of the circle in the 2D plane is

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

After expanding the equation

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

From the above equation the circle equation can be linearly separable in the feature space (x1, x2, $x1^2$, $x2^2$) with weight as (-2a, -2b, 1, 1) and the intercept as ($a^2+b^2-r^2$).

Task 4

The equation of eclipse is in the 2D plane is

- ⇨ $(x1-a)^2+d(x2-b)^2-1=0$
- ⇨ $-2cax1-2dbx2+cx1^2+dx2^2+(ca^2+db^2-1)=0$

With the feature space as ($x1$, $x2$, $x1^2$, $x2^2$, $x1x2$) the weights are (-2ca,-2db,c,d,0) and the intercept is (ca^2+db^2-1)

We prove that poly kernel of degree 2 K(u,v) = (1 + u.v)^2 is equivalent to feature space(x1, x2, x1^2, x2^2, x1 * x2)

Use u = x1, v = x2;

$K(x1,x2) = (1+x1 * x2)^2$

We expand the kernel function using 2 2D vectors(x1(x11,x12) and x2(x21,x22) to show the matrix with coefficients and matrix of terms is equivalent to those of the ellipse.

$K(x1,x2)=(1+(x11,x12).(x21,x22))^2$

$=(1+x11.x21+x12x22)^2$

$=(1+2x11x21+2x12x22+x11^2x21^2+x21^2x22^2+2x11x12x21x22)$

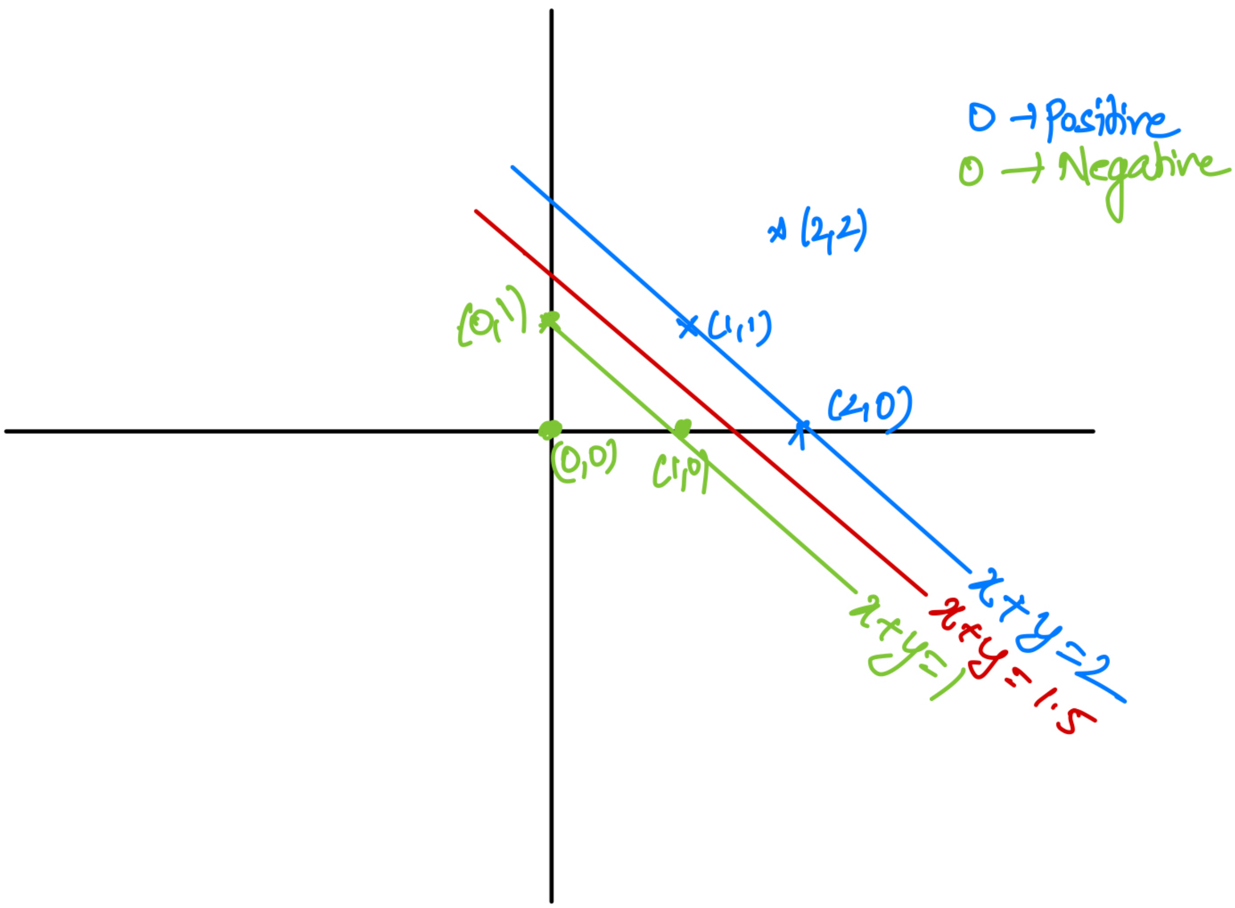
Utilizing matrix multiplication:

(1,2x11,2x12, x11², x12²,2x11x12) . (1,2x21,2x22, x21², x22²,2x21x22)^T

As we can see from the equation we get a similar feature space to what we have derived earlier for eclipse. So with this kernel we can separate any elliptical region from the rest of the plane.

TASK 5:

(a) Yes, the given training points are linearly separable.
(b)



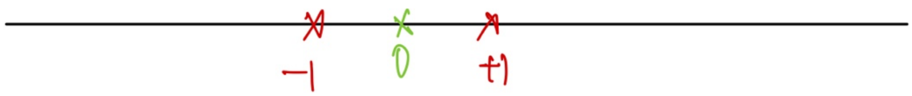
Weight vector is [1,1]^T

Support Vectors are

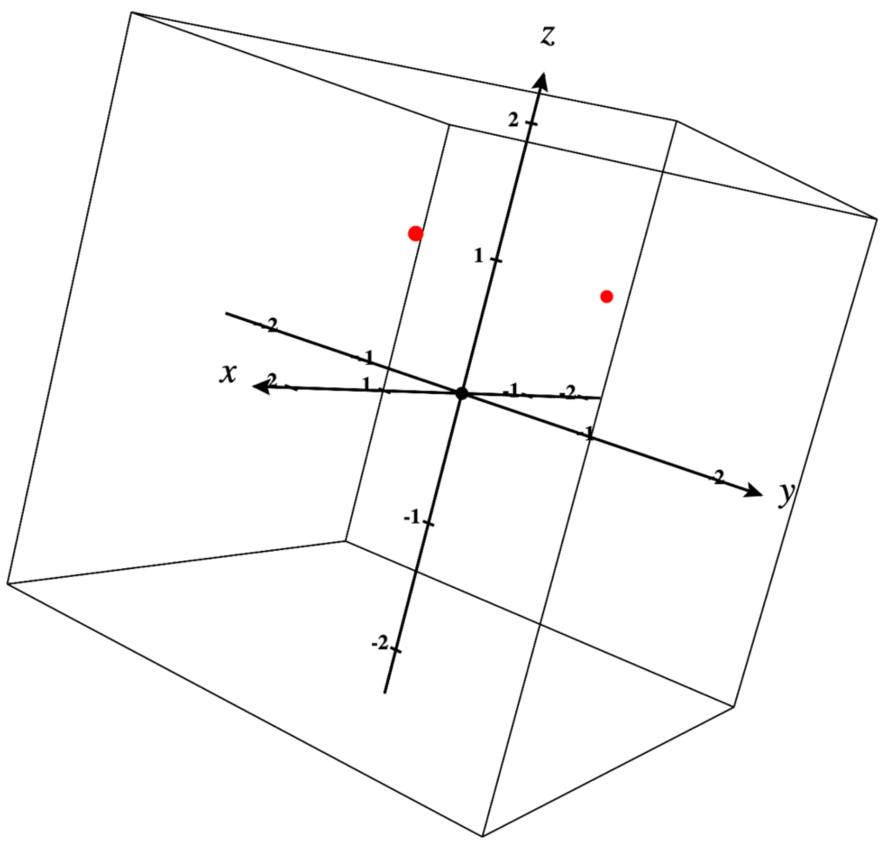
(1,0) , (0,1) for Negative Data points with line equation y = -x + 1

(1,1), (2,0) for Positive Data Points with line equation y = -x + 2

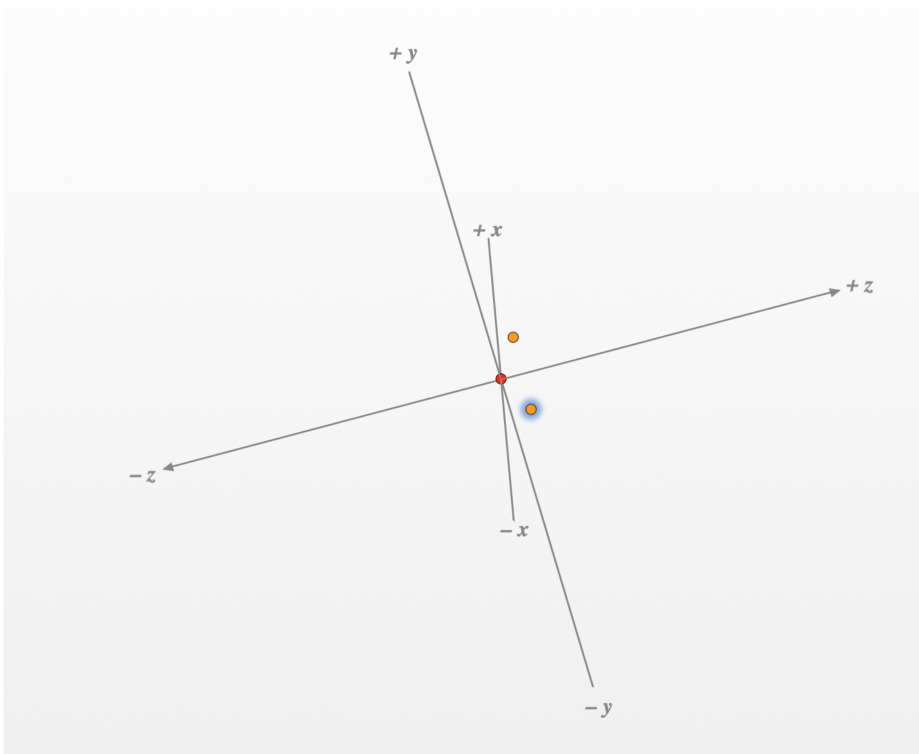
TASK 6:



Red – Negative Class
Green – Positive Class
(a) No the classes + and – are not linearly separable.
(b) Using new feature vectors f(x) = [1,sqrt(2)*x, x^2], the points are now linearly seperable.



A hyperplane parallel to XY plane and with z coordinate 0.5 will be the best hyperplane z = ½;



TASK 7:

Accuracy with 5-fold cross-validation was 84.70 %

RBF Kernel Performed better than ‘Linear’ or ‘quadratic’ Kernel.

Google Colab notebook link –

https://colab.research.google.com/drive/1Gb7rw8aswWjSpxmfbK2vl5qP1yQ-vCR7?usp=sharing