Q3 (a) maximize $f(x,y) = xy$

constraint 1 $x + y^2 \le 2$

constraint 2 $x > 0$

constraint 3 $y > 0$

~~case~~ lagrange equation for given system:

$$L = xy + \lambda_1(2 - x - y^2) + \lambda_2(-x) + \lambda_3(-y)$$

<u>case 1</u>: Assume,

$\underline{\lambda_1 = 0}$ (i.e., constraint 1 is non-binding)

$$\Rightarrow L = xy + \lambda_2(-x) + \lambda_3(-y)$$

$L_x = y - \lambda_2$ $y - \lambda_2 = 0 \Rightarrow y = \lambda_2$

$L_y = x - \lambda_3$ $x - \lambda_3 = 0 \Rightarrow x = \lambda_3$

$L_{\lambda_2} = -x$ $-x = 0 \Rightarrow x = 0$

$L_{\lambda_3} = -y$ $-y = 0 \Rightarrow y = 0$

$$\Rightarrow \boxed{x = y = \lambda_2 = \lambda_3 = 0}$$

checking constraint 1:

$x + y^2 \le 2$

$0 + 0 \le 2$

$0 \le 2$ True.

$\therefore$ KKT conditions are satisfied when

$x = y = \lambda_2 = \lambda_3 = \lambda_1 = 0$

$\therefore \boxed{xy = 0.}$

$\Rightarrow \boxed{f(x,y) = 0}$

↳ although this is not a local max.

since $f(0,0) = 0$

but $f(x,y) > 0$ at some points.

Case 2: Assume $\lambda_3 = 0$

(i.e, the constraint 3 is non binding)

$\Rightarrow L = xy + \lambda_1(2 - x - y^2) + \lambda_2(-x)$

$L_x = \cancel{y} + (-\lambda_1) - \lambda_2 = y - \lambda_1 - \lambda_2$

$L_y = x + (-2y\lambda_1) = x - 2y\lambda_1$

$L_{\lambda_1} = 2 - x - y^2$

$L_{\lambda_2} = -x$

$y - \lambda_1 - \lambda_2 = 0$

$x - 2y\lambda_1 = 0$

$2 - x - y^2 = 0$

$x = 0$

$\Rightarrow y = \lambda_1 + \lambda_2$

$\boxed{x = 0}$

$\boxed{\lambda_1 = 0}$

$x = 2y\lambda_1$

$2 - x - y^2 = 0$

$2 - 2y\lambda_1 - y^2 = 0$

$2 - y^2 = 0$

$y^2 = 2$

$\boxed{\Rightarrow y = \sqrt{2}}$

since $\lambda_1 = 0$,

It falls back to

the case 1 which already satisfies KKT.

i.e, $\boxed{f(x,y) = 0.}$

Case 3: Assume $\lambda_2 = 0$

(ie, the constraint 2 is non binding)

$\Rightarrow L = xy + \lambda_1(2 - x - y^2) + \lambda_3(-y)$

$L_x = y - \lambda_1$
$L_y = x - 2y\lambda_1 - \lambda_3$
$L_{\lambda_1} = 2 - x - y^2$
$L_{\lambda_3} = -y$

$y - \lambda_1 = 0 \Rightarrow y = \lambda_1$

$\Rightarrow x - 2y^2 - \lambda_3 = 0 \Rightarrow x = 2y^2 + \lambda_3$

$\Rightarrow 2 - x - y^2 = 0 \Rightarrow 2 - 2y^2 - \lambda_3 - y^2 = 0$

$\Rightarrow 2 - 3y^2 - \lambda_3 = 0$

$\Rightarrow 3y^2 = \lambda_3 - 2$

$y = 0$
$\lambda_1 = 0$
$\Rightarrow \lambda_3 = 0$

$y^2 = 2/3$

$\boxed{y = \sqrt{2/3}}$

$\Rightarrow 2 - x - y^2 = 0 \Rightarrow 2 - x - 2/3 = 0$

$2 - 2/3 = x$

$\boxed{\Rightarrow x = 4/3}$

checking constraint 2

$x > 0$

here $x = 4/3 > 0$ $\therefore$ KKT is satisfied

$\Rightarrow xy = 4/3 \cdot \sqrt{2/3}$ $\boxed{\Rightarrow f(x,y) = \frac{4}{3}\sqrt{\frac{2}{3}}}$

$\therefore$ The local max are $(0,0)$ & $(4/3, \sqrt{2/3})$
& the global max is at $(4/3, \sqrt{2/3})$.

(b) TRUE,

Given a linearly separable data, the margin of the decision boundary produced by SVM will always be greater than or equal to the margin of the decision boundary produced by any other hyperplane that perfectly classifies the data for the given training dataset because;
SVM is good are creating the margin that is maximal as it is also called as maximal-margin classifier. SVM has the liberty of finding support vectors which the other classifiers do not have.
Since it is a maximal-margin classifier it finds the support vectors which are at max. distance to any other training instance. No other classifier can gaurentee that it will find the maximal margin.

Q4. (a) $K(x,x') = cK^{(1)}(x,x')\ ; c>0$

pre-multiplying by z^T we get

$$z^T c . K^{(1)}(x,x') \geq 0$$

now,

post-multiplying by z we get

$$\underline{z^T c . K^{(1)}(x,x') z \geq 0}$$

∴ This satisfies & validates that this is a valid kernal since it forms a PSD kernel matrix

(b) $K(x,x') = K^{(1)}(x,x') + K^{(2)}(x,x')$

again, pre multiplying both parts of the eqn by z^T, we get

$$z^T K^{(1)}(x,x') + z^T K^{(2)}(x,x') \quad \text{— ①}$$

now,

post-multiplying both parts of the eqn ① by z, we get

$$z^T K^{(1)}(x,x')z + z^T K^{(2)}(x,x') z \geq 0$$

$$\Rightarrow z^T \left[K^{(1)}(x,x') + K^{(2)}(x,x')\right] z \geq 0$$

$$\Rightarrow \underline{z^T (K) z \geq 0}$$

since it forms a positive semi-definite kernal matrix

∴ This satisfies & validates that this is a valid kernal.

(c) $k(x,x') = f(x)\, k^{(1)}(x,x')\, f(x')$ — ①
where f is any function from R^m to R

decomposing $k^{(1)}(x,x')$ from eqn ①

$$k^{(1)}(x,x') = \phi^{(1)}(x)\, \phi^{(1)}(x') \quad — ②$$

using eqn ② in ①
∴ we get

$$f(x)\, \phi^{(1)}(x)\, \phi^{(1)}(x')\, f(x') \quad — ③$$

rearranging eqn ③ we get

$$[f(x)\, \phi^{(1)}(x)]\, [f(x')\, \phi^{(1)}(x')]$$

putting $\phi^{(2)}(x) = f(x)\, \phi^{(1)}(x)$
∴ $\phi^{(2)}(x')$ becomes $f(x')\, \phi^{(1)}(x')$

$$\Rightarrow \underline{k(x,x') = \phi^{(2)}(x)\, \phi^{(2)}(x')}$$

∴ This satisfies & validates that this is a valid kernel.

(d) $k(x, x') = k^{(1)}(x, x')\, k^{(2)}(x, x')$

$\Rightarrow k(x, x') = \phi^{(1)}(x^T)\, \phi^{(1)}(x')\, \phi^{(2)}(x^T)\, \phi^{(2)}(x')$

since $k^{(1)}$ & $k^{(2)}$ are valid kernals

by mercers theorm:

$k_1(x, x') = \phi_1(x)^T \phi_1(x') \rightarrow$ m-dimensional vector

$k_2(x, x') = \phi_2(x)^T \phi_2(x') \rightarrow$ n-dimensional vector

$\therefore k(x, x') = \sum_{i=1}^{n} \sum_{j=1}^{m} \phi_i^{(1)}(x)\, \phi_i^{(1)}(x')\, \phi_j^{(2)}(x)\, \phi_j^{(2)}(x')$

$\Rightarrow k(x, x') = [\phi_1^{(1)}(x)\, \phi_1^{(2)}(x)\, \phi_1^{(1)}(x)\, \phi_2^{(2)}(x) \cdots \phi_n^{(1)}(x)\, \phi_m^{(2)}(x)]$

$$\cdot \begin{bmatrix} \phi_1^{(1)}(x')\, \phi_1^{(2)}(x') \\ \vdots \\ \phi_n^{(1)}(x')\, \phi_m^{(2)}(x') \end{bmatrix}$$

now, we can write $\phi^{(3)} x$ as

$[\phi_1^{(1)}(x)\, \phi_1^{(2)}(x)\, \phi_1^{(1)}(x)\, \phi_2^{(2)} x \cdots \phi_n^{(1)}(x)\, \phi_m^{(2)}(x)]^T$

$$\boxed{\Rightarrow k(x, x') = \phi^{(3)}(x)^T \phi^{(3)}(x')}$$

$\hookrightarrow$ PSD kernal matrix

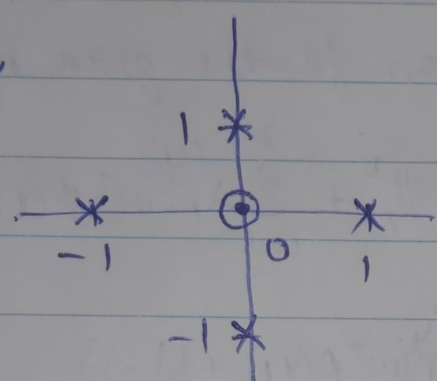
$\therefore$ This is a valid kernel.

Qs (a)

class	x
+	0
−	−1
−	+1

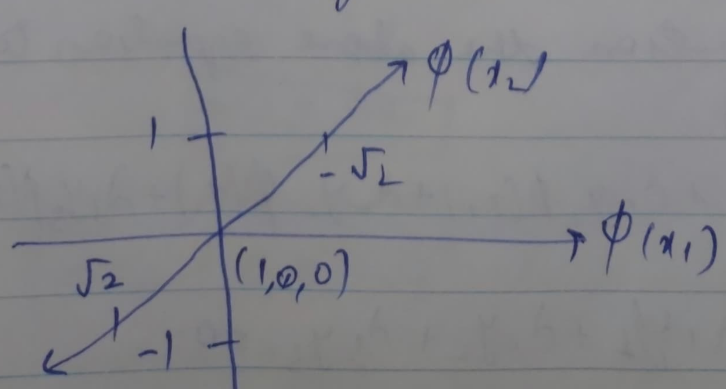
class
O → +
× → −

Example,



No, the classes are not linearly seperable.

(b) since, $\phi(x) = [1, \sqrt{2}x, x^2]^T$
we know, for +ve : $x = 0$
$\therefore \phi(x_1) = (1, 0, 0)$
for −ve : $x = -1$ & $x = 1$
$\phi(x_2) = (1, -\sqrt{2}, 1)$ $\phi(x_3) = (1, \sqrt{2}, 1)$
$\therefore$ yes the hyperplane is now linearly seperable with weight vectors $w = (0, 0, 1)$



(c) Applying Lagrange multipliers:

$$\min_{w,b} \frac{1}{2}\|w\|_2^2$$

such that,

$$y_i(w^T\phi(x_i)+b)=1, \quad i=1,2,3$$

Lagrange equation for the given system

$$L = \frac{1}{2}\|w\|_2^2 + \sum_{i=1}^{3}\lambda_i(y_i(w^T\phi(x_i)+b)-1)$$

$$\Rightarrow L = \frac{1}{2}\|w\|_2^2 + \lambda_1(y_1(w^T\phi(x_1)+b)-1)$$
$$+ \lambda_2(y_2(w^T\phi(x_2)+b)-1)$$
$$+ \lambda_3(y_3(w^T\phi(x_3)+b)-1)$$

now,

$$L_w = \frac{\partial L}{\partial w} = w + \lambda_1 y_1\phi(x_1)$$
$$+ \lambda_2 y_2 \phi(x_2)$$
$$+ \lambda_3 y_3 \phi(x_3)$$

$$L_b = \frac{\partial L}{\partial b} = \lambda_1 y_1 + \lambda_2 y_2 + \lambda_3 y_3$$

now, equating the above equation to 0 we get

$$w + \lambda_1 y_1 \phi(x_1) + \lambda_2 y_2 \phi(x_2) + \lambda_3 y_3 \phi(x_3) = 0$$

&

$$\lambda_1 y_1 + \lambda_2 y_2 + \lambda_3 y_3 = 0$$

we know that $\phi(x) = (1, \sqrt{2}x, x^2)^T$

$\therefore$ we get

$\Rightarrow$ $w_1 + \lambda_1 - \lambda_2 - \lambda_3 = 0$ — 1

$w_2 + \sqrt{2}\lambda_2 - \sqrt{2}\lambda_3 = 0$ — 2

$w_3 - \lambda_2 - \lambda_3 = 0$ — 3

$\lambda_1 - \lambda_2 - \lambda_3 = 0$ — 4

using 1 & 4 we get

$\Rightarrow$ $w_1 + 0 = 0$

$\boxed{\Rightarrow w_1 = 0}$

if $w_1 = 0$,

$\Rightarrow$ $\boxed{b = 1}$

also,

$-\sqrt{2}w_2 + w_3 + b = -1$ — 5

$\sqrt{2}w_2 + w_3 + b = -1$ — 6

adding in 5 & 6 we get

$2w_3 + 2b = -2$

$2w_3 + 2 = -2$

$2w_3 = -4$

$\boxed{w_3 = -2}$

$\Rightarrow \sqrt{2}w_2 = -1 + 1$

$w_2 = \frac{0}{\sqrt{2}}$

$\boxed{w_2 = 0}$

$\therefore$ $\boxed{\hat{w} = (0, 0, -2)^T \text{ \& } b = 1 \text{ \& margin} = 1/2}$

(d) Now, for the given equation

$$y_i(w^T\phi(x_i)+b) \geq \rho, \quad i=1,2,3$$

Every equation is similar to previous question

so,

$$L = \frac{1}{2}\|w\|_2^2 + \lambda_1 y_1(w^T\phi(x_1)+b) - \rho)$$
$$+ \lambda_2 y_2(w^T\phi(x_2)+b) - \rho)$$
$$+ \lambda_3 y_3(w^T\phi(x_3)+b) - \rho)$$

Ily we get the following equation

$$w_1 + \lambda_1 - \lambda_2 - \lambda_3 = 0 \quad -(1)$$
$$w_2 + \sqrt{2}\lambda_2 - \sqrt{2}\lambda_3 = 0 \quad -(2)$$
$$w_3 - \lambda_2 - \lambda_3 = 0 \quad -(3)$$
$$\lambda_1 - \lambda_2 - \lambda_3 = 0 \quad -(4)$$

using (1) & (4) we get

$$w_1 + 0 = 0$$
$$w_1 = 0$$

if $w_1 = 0$

$$\boxed{b = \rho}$$

also,
$$-\sqrt{2}w_2 + w_3 + b = -1 \quad -(5)$$
$$+\sqrt{2}w_2 + w_3 + b = -1 \quad -(6)$$

adding (5) & (6) we get $\rightarrow \boxed{w_3 = -2\rho}$

& $\boxed{w_2 = 0}$

$\therefore \boxed{b = \rho \;\;\&\;\; \hat{w} = (0, 0, -2\rho)^T}$

Now, $w^T x + b = 0$

$x : -2\rho x_3 + \rho = 0$

also for the previous question

$x : -2x_3 + 1 = 0$

$\therefore$ The solution remains same even if the constraint to ρ.

(e) Lets assume the distribution of dataset changes. we know, any linear or convex optimization, that deals with constraints to maximize or minimize a given objective function, say $f(x)$, depends on a particular parameter of optimization.

given any f, our pardigm or the authentic structure allows us to substitute an feature vector with its scaled counterpart (ie, if $\vec{x}$ is a vector. $k\vec{x}$ is scaled version of $\vec{x}$ where $k \in R$), then the optimization of $f(x)$ becomes,

$$f(k) \cdot f(\vec{x})$$

we know

$f(k)$ is independent of parameter of optimization or the variable.

$\therefore f(k)$ is a constant.

$\therefore$ if the goal is min $f(k) - f(x)$

$$= f(k) \min_{\theta} f(x)$$

which is the same optimization problem as the constant $f(k)$ do not play any role.

$\therefore$ This ~~dataset~~ is true for any dataset.