Q13)(a)

$$\max_{x,y} f(x,y) = xy \quad \text{s.t.} \quad x+y^2 \le 2, \; x,y > 0 \quad \text{using KKT conditions}$$

sol

Converting to lagrange eqn :-

$$L(x,y,\lambda) = xy - \lambda(x+y^2-2)$$

Applying KKT conditions :

1. $\nabla_x L = 0 \Rightarrow y - \lambda = 0 \Rightarrow \boxed{y=\lambda}$ (No feasible descent) (1a)

$\nabla_y L = 0 \Rightarrow x - 2\lambda y = 0 \Rightarrow x = 2\lambda y$ or $\boxed{x = 2y^2}$ (1b)

2. $\lambda(x+y^2-2) = 0$ (Complementary Slackness)

3. $x+y^2-2 \le 0$ (Feasible Constraints)

4. $\lambda \ge 0$ (Positive lagrange Multiplier)

Case 1 $\lambda = 0$.

Subs. in (1a), (1b),

$y = \lambda = 0$.

$x = 2y^2 = 0$.

but $x, y > 0$ given

$\Rightarrow$ Invalid case.

Case 2: $\lambda > 0$ ($\lambda \ne 0$).

Subs in 2

$x+y^2-2=0$.

Subs (1b) here

$2y^2+y^2-2=0 \Rightarrow y = \sqrt{\frac{2}{3}}$

$x = 4/3$.

$\Rightarrow \lambda = \sqrt{2/3} > 0$

so $x, y > 0$ and $\lambda \geq 0$ is satisfied

Hence we get optimal values of $(x, y) = \left(\frac{4}{3}, \sqrt{\frac{2}{3}}\right)$

were all 4 KKT conditions are satisfied.

Q3(b).

Ans. True

SVM draws a hyperplane which maximises the min possible geometric margin for all pts wrt the decision boundary, i.e. to draw a decision boundary that maximises the dist b/w the closest set of pts of the diff classes (called support vectors.)

Q(4)

(a) $K(x, x') = cK^1(x, x')$ where $c \geq 0$.

$K^{(1)}$ is a valid kernel so $\forall z$ $z \in \mathbb{R}^n$, $z^T K^{(1)} z \geq 0$

Now $K(x, x') = cK^1(x, x')$

$= c \cdot z^T K^{(1)} z \geq 0 \quad (\because c \geq 0 \ \& \ z^T K^{(1)} z \geq 0)$

$$\Rightarrow z^T cK^{(1)} z \geq 0$$

$\Rightarrow cK^{(1)}$ or $K(x, x')$ is a valid Kernel $\because$ it satisfies the positive semi-def. matrix property of mercer's theorem.

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

$\because K^{(1)}$ and $K^{(2)}$ are valid kernels

$\therefore z^T K^{(1)} z \geq 0$ and $z^T K^{(2)} z \geq 0 \quad \forall z \in R^n$

$$\Rightarrow z^T K^{(1)} z + z^T K^{(2)} z \geq 0$$

$$\Rightarrow z^T (K^{(1)} + K^{(2)}) z \geq 0$$

$\therefore (K^{(1)} + K^{(2)})$ follows property of +ve semi-definite matrix.

$\therefore K = K^{(1)} + K^{(2)}$ is valid Kernel.

(c) $$K(x, x') = f(x) K^{(1)}(x, x) \cdot f(x') \quad f: R^m \rightarrow R$$

f maps an n dimension vector $x \langle x_1 \ldots x_m \rangle$ into a 1 dim vector $f(x)$, i.e. a scalar value.

Let $f(x) = c_1$ and $f(x') = c_2$.

$K^{(1)}$ is a valid kernel

$\Rightarrow z^T K^{(1)} z \geq 0 \quad \forall z \in R^n$

$\Rightarrow \quad c_1 c_2 z^T K^{(1)} z \geq 0$ where $c_1, c_2 \geq 0$

$\Rightarrow \quad z^T (c_1 K^{(1)} c_2) z \geq 0$

$\Rightarrow \quad z^T \left(f(x) \cdot K^{(1)} \cdot f(x') \right) z \geq 0$

$\Rightarrow \quad f(x) \cdot K^{(1)} f(x')$ follows +ve semi-def matrix property

$\Rightarrow \quad K = f(x) \cdot K^{(1)} f(x)$ is a valid kernel.

(d) $\quad K(x, x') = K^{(1)}(x, x') \cdot K^{(2)}(x, x')$

K^1 and K^2 are valid kernels.

$\therefore$ they corresponds to remapping of i/p to new feature space.
i.e. $K(x, y) = \sum_i \phi_i(x) \phi_i(y)$ for some large (may be ∞) set of basis functions.

So $K^{(1)}(x, x') = \sum_i \phi_i^{(1)}(x) \phi_i^{(1)}(x')$

$K^{(2)}(x, x') = \sum_j \phi_j^{(2)}(x) \cdot \phi_j^{(2)}(x')$

$K^{(1)}(x, x') \cdot K^{(2)}(x, x')$

$$= \sum_i \phi_i^{(1)}(x) \cdot \phi_i^{(1)}(x') \cdot \sum_j \phi_j^{(2)}(x) \cdot \phi_j^{(2)}(x')$$

$$= \sum_i \sum_j \phi_i^{(1)}(x) \cdot \phi_j^{(2)}(x) \; \phi_i^{(1)}(x') \cdot \phi_j^{(2)}(x')$$

$\because$ each ϕ function outputs a scalar

$\therefore$ we can define $\phi_k(x) = \phi_i^{(1)}(x) \cdot \phi_j^{(2)}(x)$

similarly for $\phi_k(x')$.

$\therefore$ we can finally write

$$K^{(1)}(x,x') \cdot K^{(2)}(x,x') = \Sigma_k \phi_k(x) \cdot \phi_k(x')$$

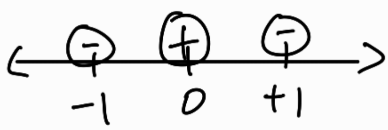
So the product of two kernels creates a function with same invariants which again correspond to remapping of i/p to new feature space (can be ∞)

$K(x,x') = K^{(1)}(x,x') \cdot K^{(2)}(x,x')$ is a valid Kernel.

Q5)

(a)

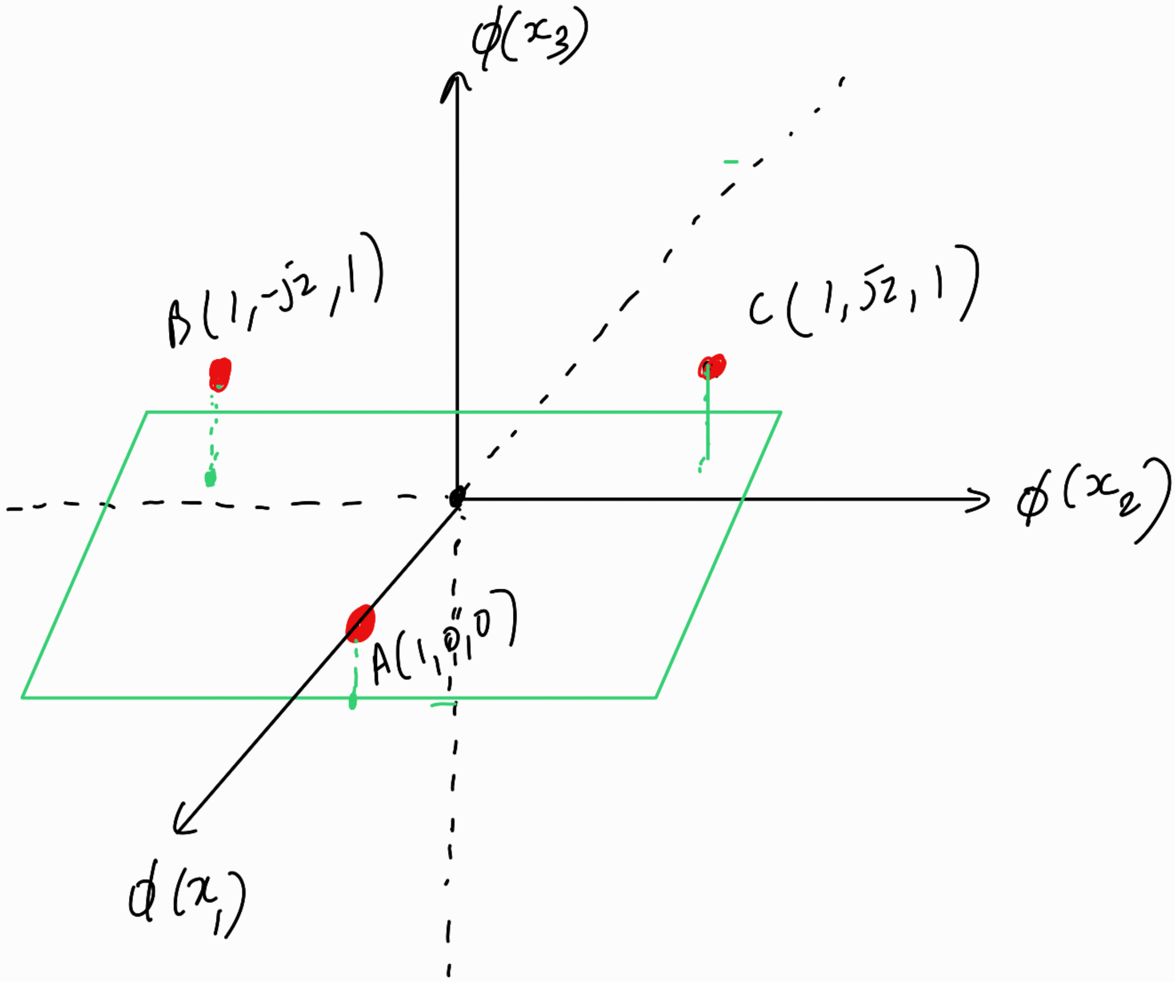
class	x
+	0
−	−1
−	+1



Classes not linearly seperable in 1D.

(b) $\phi(0) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ $\phi(-1) = \begin{bmatrix} 1 \\ -\sqrt{2} \\ 1 \end{bmatrix}$ $\phi(1) = \begin{bmatrix} 1 \\ \sqrt{2} \\ 1 \end{bmatrix}$

The pts are mapped to $(1,0,0)$, $(1,-\sqrt{2},1)$ and $(1,\sqrt{2},1)$ in 3D vector space.



A separating hyperplane is given by the wt vector $(0,0,1)$ in the new 3D space.

(c)

$$\min_{w,b} \frac{1}{2}\|w\|_2^2 \quad \text{s.t.} \quad y_i(w^T\phi(x_i) + b) \geq 1 \quad \underline{i=1,2,3.}$$

optimisation with inequality constraints is done using KKT conditions, which is generalisation of lagrange multipliers.

In this problem, we have 3 vectors in 3D space & all of them are support vectors; hence all 3 constraints hold with equality.

∴ applying method of lagrange multipliers to following problem is same as solving above

$$\left[\begin{array}{l} \min_{w,b} \ \frac{1}{2}\|w\|_2^2 \quad s.t. \\ y_i\left(w^T\phi(x_i)+b\right) = 1 \qquad i=1,2,3 \end{array}\right]$$

$$L(w, \lambda_1, \lambda_2, \lambda_3) = \frac{1}{2}\|w\|_2^2 + \sum_{i=1}^{3} \lambda_i \left(y_i\left(w^T\phi(x_i)+b\right)-1\right)$$

Applying lagrange conditions:

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w + \lambda_i y_i \phi(x_i) = 0 \quad -① \quad /$$

$$\frac{\partial L}{\partial b} = 0 \quad \sum_{i=1}^{3} \lambda_i y_i = 0 \quad -② \quad /$$

Using the data points $\phi(x_i)$, we get foll. eqns from above

$$\left.\begin{array}{l} w_1 + \lambda_1 - \lambda_2 - \lambda_3 = 0 \\ w_2 + \sqrt{2}\lambda_2 - \sqrt{2}\lambda_3 = 0 \\ w_3 - \lambda_2 - \lambda_3 = 0 \end{array}\right\} \quad w_1 = 0$$

$\lambda_1 - \lambda_2 - \lambda_3 = 0$)

Putting these values in equality constraints in optimisation constraints,

$$\left.\begin{aligned} b &= 1 \\ -\sqrt{2}w_2 + w_3 + b &= -1 \\ \sqrt{2}w_2 + w_3 + b &= -1 \end{aligned}\right\} \quad w_2 = 0, \; w_3 = -2$$

$\therefore$ optimal wts $= (0, 0, -2)^T$ and $b = 1$

Margin is $\dfrac{1}{\|w\|_2} = \dfrac{1}{\sqrt{0^2+0^2+(-2)^2}} = \dfrac{1}{2}$.

(d)

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

changing constraints only changes the last 3 eqn s,

we get $b = \rho$, $\vec{w} = (0, 0, -2\rho)^T$

hyperplane described by eqn $w^T x + b = 0$ remains same

$$\{x : -2\rho x_3 + \rho = 0\} \equiv \{x : -2x_3 + 1 = 0\}$$

Hence we have the same classifier in both cases, which assign class label + if $w^T x + b \geq 0$ and assign class label - otherwise.

(e) This is true for any dataset & it follows from homogenity of optimisation problem.

for constraints $y_i(w^T \phi(x_i) + b) \geq \rho$, we can define new weight vectors $w' = w/\rho$ and $b' = b/\rho$.

so that the constraints in new variable are $y_i(w'^T \phi(x_i) + b') \geq 1$.

And equivalently optimise the following

$$\min_{w', b'} \frac{1}{2} \rho^2 \|w'\|_2^2 \text{ s.t.}$$

$$y_i(w'^T \phi(x_i) + b') \geq 1 \quad i = 1, 2, 3.$$

Since ρ^2 is a constant, multiplying OF $\|w\|_2^2$ it does'nt change optimal value & the 2 solutions describe same classifier. $w^T x + b \geq 0 \equiv \rho w^T x + \rho b \geq 0$.