Q3) a) $f(x,y) = xy$ s.t. $x+y^2 \leq 2$ and
$x, y > 0$

The KKT conditions can be written as follows —

① converting to lagrange equation —

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

α is the lagrange constant associated with the inequality $x+y^2 \leq 2$
which can be written as — $x+y^2-2 \leq 0$

differentiating w.r.t x, y.

$$\frac{\partial L}{\partial x} = y - \alpha = 0$$
$$\underline{y = \alpha} \quad — \quad (a)$$

$$\frac{\partial L}{\partial y} = x - \alpha(2y) = 0$$
$$\underline{x = 2\alpha y} \quad — \quad (b)$$

② $\alpha \geq 0$ ③ $x+y^2-2 \leq 0$

④ $\alpha(x+y^2-2) = 0$ ⑤ $x > 0$
$y > 0$

using above conditions for KKT and solving for α, x and y —

substituting $y = \alpha$ (a) in (b) —

$$x = 2\alpha \cdot \alpha$$
$$x = 2\alpha^2$$

substituting $y = \alpha$ and $x = 2\alpha^2$ in eqn (4)

$$\alpha(2\alpha^2 + \alpha^2 - 2) = 0$$
$$\Rightarrow 2\alpha^3 + \alpha^3 - 2\alpha = 0$$
$$\Rightarrow 2\alpha^3 + \alpha^3 = 2\alpha$$
$$\Rightarrow 2\alpha^2 + \alpha^2 = 2$$
$$\Rightarrow 3\alpha^2 = 2$$
$$\Rightarrow \alpha^2 = \frac{2}{3}$$
$$\Rightarrow \alpha = \pm\sqrt{2/3}$$

neglecting $-\sqrt{2/3}$ according to condn (2)

Thus $\alpha = \sqrt{\frac{2}{3}}$

now $y = \alpha = \sqrt{2/3}$ [> 0 as in eqn (5)]

$x = 2\alpha^2 = 2 \times \frac{2}{3} = \frac{4}{3}$ [> 0 as in eqn (6)]

substituting these values of x & y in eqn (3) —

$$x + y^2 - 2 \leq 0$$
$$\Rightarrow \frac{4}{3} + \frac{2}{3} - 2 \leq 0$$
$$\Rightarrow \frac{6}{3} - 2 \leq 0$$
$$\Rightarrow 2 - 2 \leq 0$$
$$\Rightarrow 0 \leq 0 \quad \text{True}$$

Thus we have global maxima at $\left(4/3, \sqrt{2/3}\right)$ and

$$\max f(x,y) = \frac{4}{3} \times \sqrt{\frac{2}{3}} = \sqrt{\frac{2}{3}} \cdot \frac{4}{3}$$

Q3) b) The above statement is true.

SVM always maximises the minimum functional / geometric margin. Thus it will always produce a decision boundary greater than equal to decision boundary of any other hyperplane, which will not always maximise the minimum functional margin. Even if it does the decision boundary will be equal to that of SVM and for the rest of the cases it will be lesser.

classmate

Date
Page

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

~~$k^{(1)}$~~

Let $k^{(1)}(x, x') = \langle \phi^{(1)}(x), \phi^{(1)}(x') \rangle$
where $\phi^{(1)}$ = feature map.

Let $\phi(x) = \sqrt{c}\,\phi(x)$.

So for all x and x' →

$$K(x, x') = c\,k^{(1)}(x, x') =$$

$$c \langle \phi^{(1)}(x), \phi^{(1)}(x') \rangle =$$

$$\langle \sqrt{c}\,\phi(x), \sqrt{c}\,\phi(x') \rangle$$

Thus $K(x, x')$ is a valid kernel corresponding to the feature map ~~of~~ ϕ.

Q4) b) $K(x, x') = k^{(1)}(x, x') + k^{(2)}(x, x')$

Let $\phi^{(1)}(x) = (\phi_1^{(1)}(x), \dots \phi_{m_1}^{(1)}(x))$
$\phi^{(2)}(x) = (\phi_1^{(2)}(x), \dots \phi_{m_2}^{(2)}(x))$

where $\phi^{(1)}(x)$ and $\phi^{(2)}(x)$ are the feature maps of $k^{(1)}$ and $k^{(2)}$.

$\phi(x)$, feature map of $K(x, x')$ = concatinating the feature maps ~~of~~ $\phi^{(1)}(x)$ and $\phi^{(2)}(x)$

$$\phi(x) = (\phi_1^{(1)}(x), \dots \phi_{m_1}^{(1)}(x), \phi_1^{(2)}(x), \dots \phi_{m_2}^{(2)}(x))$$

Thus,

$$\phi(x) \cdot \phi(x') = \phi^{(1)}(x) \cdot \phi^{(1)}(x') + \phi^{(2)}(x) \cdot \phi^{(2)}(x')$$

proved.

Q4) c) $K(x, x') = f(x)\, K^{(1)}(x, x')\, f(x')$

as $K^{(1)}(x, x')$ is a valid kernel, ~~function~~ thus $\phi_1(x)^T \phi_1(x')$ will give a positive semidefinite matrix.

now,

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

$$\Rightarrow f(x) \langle \phi(x), \phi(x') \rangle f(x')$$

$$\Rightarrow \langle f(x)\, \phi(x), f(x')\, \phi(x') \rangle$$

$$\Rightarrow \langle \phi'(x), \phi'(x') \rangle$$

$\Rightarrow K(x, x')$ which is a valid kernel function.

proved

Date
Page

a4) d) $K(x, x') = k^{(1)}(x, x')\, k^{(2)}(x, x')$

let $k^{(1)}(x, x') = \langle \phi^{(1)}(x), \phi^{(1)}(x') \rangle$
and $k^{(2)}(x, x') = \langle \phi^{(2)}(x), \phi^{(2)}(x') \rangle$

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

$$= \langle \phi^{(1)}(x), \phi^{(1)}(x') \rangle \cdot \langle \phi^{(2)}(x), \phi^{(2)}(x') \rangle$$

$$= \sum_{m,n=1}^{k} \left(\phi_m^{(1)}(x)\, \phi_n^{(2)}(x) \right) \cdot \left(\phi_m^{(1)}(x')\, \phi_n^{(2)}(x') \right)$$

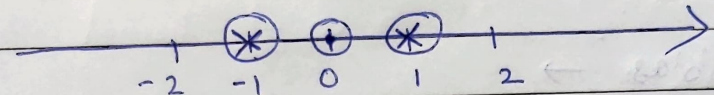
where k = k dimensional vectors.
ie x and x' are k dimensional vectors

$$= \langle \phi(x), \phi(x') \rangle$$

where $\phi_{m,n}(x) = \phi_m^{(1)}(x) \cdot \phi_n^{(2)}(x')$ ie a ~~$k^2 \times 1$ feature~~ $k^2 \times 1$ feature map.

Thus $k(x, x')$ is a valid kernel with feature map ϕ.

Q5) a the classes can be roughly represented in 1D as —



class

⊙ → +

⊗ → −

no, the classes are not linearly separable as seen from above

Q5) b) new feature vector $\phi(x) = \begin{bmatrix} 1 \\ \sqrt{2}x \\ x^2 \end{bmatrix}$

mapping the above 3 points —

Date
Page

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

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

Yes the classes are now linearly separable in 3 dimensional plane.

Finding the separating hyperplane —

Let the 3 points be denoted as →

$$s_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad s_3 = \begin{bmatrix} 1 \\ \sqrt{2} \\ 1 \end{bmatrix} \quad \text{and}$$

$$s_2 = \begin{bmatrix} 1 \\ -\sqrt{2} \\ 1 \end{bmatrix}$$

adding bias →

$$\bar{s_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \qquad \bar{s_2} = \begin{bmatrix} 1 \\ \sqrt{2} \\ 1 \\ 1 \end{bmatrix} \qquad \bar{s_3} = \begin{bmatrix} 1 \\ \sqrt{2} \\ 1 \\ 1 \end{bmatrix}$$

putting the above in the below equation —

$$\alpha_1 \bar{s_1}\bar{s_1} + \alpha_2 \bar{s_1}\bar{s_2} + \alpha_3 \bar{s_1}\bar{s_3} = 1$$
$$\alpha_1 \bar{s_1}\bar{s_2} + \alpha_2 \bar{s_2}\bar{s_2} + \alpha_3 \bar{s_3}\bar{s_2} = -1$$
$$\alpha_1 \bar{s_1}\bar{s_3} + \alpha_2 \bar{s_2}\bar{s_3} + \alpha_3 \bar{s_3}\bar{s_3} = -1$$

$\bar{s_1}$ is common so equating its respective class ie +ve i.e s_1 is present in ~~positive~~ +ve labelled class.

$$\alpha_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ -\sqrt{2} \\ 1 \\ 1 \end{pmatrix} +$$

$$\alpha_3 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ \sqrt{2} \\ 1 \\ 1 \end{pmatrix} = 1$$

$$\alpha_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ -\sqrt{2} \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ -\sqrt{2} \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ -\sqrt{2} \\ 1 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 1 \\ -\sqrt{2} \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ \sqrt{2} \\ 1 \\ 1 \end{pmatrix}$$

$$= -1$$

$$\alpha_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ \sqrt{2} \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ \sqrt{2} \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ -\sqrt{2} \\ 1 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 1 \\ -\sqrt{2} \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ \sqrt{2} \\ 1 \\ 1 \end{pmatrix}$$

$$= -1$$

we get $\rightarrow$

$$2\alpha_1 + 2\alpha_2 + 2\alpha_3 = 1$$
$$2\alpha_1 + 5\alpha_2 + \alpha_3 = -1$$
$$2\alpha_1 + \alpha_2 + 5\alpha_3 = -1$$

solving the above equation we get,

$$\alpha_1 = \frac{5}{2} \qquad \alpha_2 = -1 \qquad \alpha_3 = -1$$

we know,

$$\bar{w} = \sum \alpha_i \bar{s_i}$$

Date
Page

2) $\frac{\sqrt{2}}{2}\begin{pmatrix}1\\0\\0\\1\end{pmatrix} - 1\begin{pmatrix}1\\-\sqrt{2}\\1\\1\end{pmatrix} - 1\begin{pmatrix}1\\\sqrt{2}\\1\\1\end{pmatrix}$

2) $\begin{pmatrix}1/2\\0\\-2\\1/2\end{pmatrix}$

thus $w = \begin{pmatrix}\frac{1}{2}\\0\\-2\end{pmatrix}$ and $b = \frac{1}{2}$.

ans.

Date
Page

Q5) c) The constraint above is an inequality constraint. Such problems can be solved by KKT. But as the points are support vectors thus the constraint can be changed to equality (since support vectors have functional margin $= 1$) and the problem can be solved by lagrange multiplier.

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

$$\forall i = 1, 2, 3$$

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

where λ = Lagrange constant associated with equality constraint and $y^i(w^T \phi(x^i) + b) = 1$ can be written as —

$$y^i (w^T \phi(x^i) + b) - 1 = 0.$$

now $w = (w_1 \quad w_2 \quad w_3)$ and using $\phi(x_i)$ data points we get —

$\frac{\partial L}{\partial w_1}$ $\quad w_1 + \lambda_1 - \lambda_2 - \lambda_3 = 0.$ (1)

$\frac{\partial L}{\partial w_2}$ $\quad w_2 + \sqrt{2}\lambda_2 - \sqrt{2}\lambda_3 = 0$ (2)

$\frac{\partial L}{\partial w_3}$ $\quad w_3 - \lambda_2 - \lambda_3 = 0$ (3)

$\frac{\partial L}{\partial b}$ $\quad \lambda_1 - \lambda_2 - \lambda_3 = 0.$ (4)

using ① and ④

$$\underline{\underline{w_1 = 0}}.$$

so in ①, we get $\underline{\underline{b = 1}}$.

~~eqn ② ③ —~~

~~according to conditions.~~

~~λ_2 (~~

we also get —

$$-\sqrt{2}\, w_2 + w_3 + b = -1 \quad ⓐ$$

$$\sqrt{2}\, w_2 + w_3 + b = -1 \quad ⓑ$$

we get these equations from the constraints —

$$+\lambda_2 \left[-(w_1 + \sqrt{2}\, w_2 + w_3 + b) - 1 \right] = 0$$

(as $\lambda_2 \geq 0$) if $\lambda_2 = 0$ then

$$-(w_1 + \sqrt{2}\, w_2 + w_3 + b) - 1 = 0$$

as $w_1 = 0$ so,

$$-\sqrt{2}\, w_2 + w_3 + b = -1$$

similarly,

$$\lambda_3 \left[-(w_1 + \sqrt{2}\, w_2 + w_3 + b) - 1 \right] = 0$$

(as $\lambda_3 \geq 0$) if $\lambda_3 = 0$ then

$$-(w_1 + \sqrt{2}\, w_2 + w_3 + b) - 1 = 0$$

as $w_1 = 0$ so

$$+\sqrt{2}\, w_2 + w_3 + b = -1$$

solving (a) & (b) we get —

$\underline{w_2 = 0}$ $\underline{w_3 = -2}$

Thus $w = \hat{w} = (0, 0, -2)^T$ & $b = 1$

and margin —

$w + b = 1$

2) $2 \cdot x_1 = 1$

2) $x_1 = \frac{1}{2}$ so margin $= \frac{1}{2}$.

Q5) d) $y_i(w^T \phi(x_i) + b) \geq \rho \quad , i = 1,2,3$

can be written as,

$$y_i(w^T \phi(x_i) + b) - \rho \geq 0$$

dividing by ρ,

$$y_i \left(\frac{w^T}{\rho} \phi(x_i) + \frac{b}{\rho} \right) - 1 \geq 0$$

we can rescale the hyperplane, by the scale invariant property of ~~functional~~ geometric margin.

thus $w'^T = \frac{w^T}{\rho}$ & $b' = \frac{b}{\rho}$

and $w'^T = (w_1, w_2, w_3)$.

we can find out the values similar to what we did in Q5) c)

Date
Page

Q5) e) Yes, the answer in Q5) d is true. It is because of the fact that functional margin is scale variant. but geometric margin is scale in variant

Let us have a functional margin —

$$\rho = y(w^T \phi(x_i) + b)$$

normalising we get, $\frac{\rho}{\|w\|}$ (geometric margin)

$$\min_{\rho, w, b} \frac{\rho}{\|w\|} \quad s.t \quad y_i(w^T \phi(x_i) + b) \geq \rho$$

$$\forall i = 1, 2, 3$$

— optimisation function.

dividing by ρ —

$$\min_{\rho, w, b} \frac{1}{\|w\|} \quad s.t \quad y_i(w^T \phi(x_i) + b) \geq 1$$

$$\forall i \geq 1$$

Thus it becomes independent of ρ.