__/__/____

Ritisha Gupta ; MT22056 $\boxed{\text{ML-ASS3}}$

Ques 3a) $\max f(x,y) = x$ s/t $x + y^2 \leq 2$

$x, y > 0$

S-1 Lagrange eq^n $L(x,y,\lambda_1) = f(x,y) - \lambda g_i(x,y)$

$= xy - \lambda_1 (x + y^2 - 2) = 0$

The KKT conditions are -

① Lagrange derivative

$\frac{\partial L}{\partial x} = 0 \Rightarrow y - \lambda_1 = 0$ — $\boxed{1.a}$

$\frac{\partial L}{\partial y} = 0 \Rightarrow x - \lambda_1 (2y) = 0$ — $\boxed{1.b}$

② Complementarity

$\lambda_i g_i(x) = 0$ $\lambda_1 (x + y^2 - 2) = 0$ — ②

③ Feasibity.

$g_i(x) \leq 0, \lambda_i \geq 0$ $x + y^2 - 2 < 0$ — (3.a)

$\lambda \geq 0$ — (3.b)

case 1: $\lambda_1 = 0$

eq 1.a $\Rightarrow$ $y = 0$

eq 1.b $\Rightarrow$ $x = 0$.

but this doesn't satisfy our original constraint $x, y > 0$ so this is not the valid case.

case 2. $\lambda_1 \neq 0$

from eq 2 $\Rightarrow x + y^2 - 2 = 0$ (should be true) — ④

from eq 1.a $\Rightarrow$ $y = \lambda_1$

from eq 1.b $\Rightarrow$ $x = 2y\lambda_1$

$x = 2\lambda_1^2$

substitute value of x & y in eq ④

$$x+y^2-2=0$$
$$2\lambda_1^2+\lambda_1^2-2=0$$
$$3\lambda_1^2=2$$
$$\lambda_1^2=\frac{2}{3}$$

$$\lambda_1=\pm\sqrt{\frac{2}{3}}$$

but from eq 3b $\Rightarrow$ $\lambda_1 \geq 0$

so $\boxed{\lambda_1=\sqrt{\frac{2}{3}}}$

Finding x & y

$$x=2\lambda_1^2=2\left(\sqrt{\frac{2}{3}}\right)^2=\frac{4}{3}$$

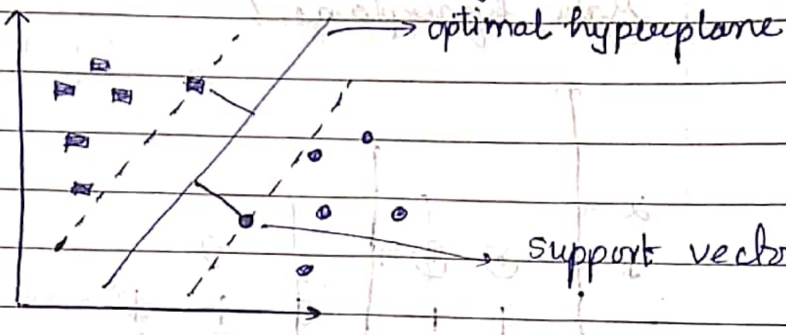
$$y=\lambda_1=\sqrt{\frac{2}{3}}$$

so optimal value of $(x,y)=\left(\frac{4}{3},\sqrt{\frac{2}{3}}\right)$

& $\lambda_1=\sqrt{\frac{2}{3}}$ — ans

Max value of $f(x,y)=xy=\frac{4}{3}\cdot\sqrt{\frac{2}{3}}=\frac{4\sqrt{2}}{9}$

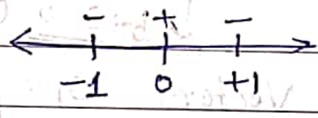
Ques 3b) The statement "Given a linearly seperable data, the margin of decision boundary produced by SVM will always be greater than or equal to the margin of decn boundry produced by any other hyperplane" is TRUE because the objective of SVM is to identify a line that maximises the minimum of functional margin. We find the points closest to the line from both the classes & these points are k/a support vectors. SVM is also k/a Optimal Margin Classifier.



Ques 5) a)

class	x
+	0
−	−1
−	+1

Plotting on 1D line



We can't draw a linear boundary b/w these 2 classes & hence they are not seperable in 1D.

b) $\phi(x) = [1, \sqrt{2}x, x^2]^T$ $\phi(0)$ $\phi(-1)$ $\phi(1)$

In 3D the new points are $(1, 0, 0)$ $(1, -\sqrt{2}, 1)$ $(1, \sqrt{2}, 1)$.

These points are seperable in 3D.

Next task is to find a hyperplane. For that we'll Assume all these 3 points $\phi(0)$, $\phi(-1)$, $\phi(1)$ as support vectors. To give vector representation we'll add a bias i.e. augment 1 to each of the point.

$$\bar{S_1} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad \bar{S_2} = \begin{pmatrix} 1 \\ -\sqrt{2} \\ 1 \\ 1 \end{pmatrix} \quad \bar{S_3} = \begin{pmatrix} 1 \\ \sqrt{2} \\ 1 \\ 1 \end{pmatrix}$$

$\alpha_1 \bar{S_1}\bar{S_1} + \alpha_2 \bar{S_2}\bar{S_1} + \alpha_3 \bar{S_3}\bar{S_1} = +1 \rightarrow$ (+ve class)

$\alpha_1 \bar{S_1}\bar{S_2} + \alpha_2 \bar{S_2}\bar{S_2} + \alpha_3 \bar{S_3}\bar{S_2} = -1$

$\alpha_1 \bar{S_1}\bar{S_3} + \alpha_2 \bar{S_2}\bar{S_3} + \alpha_3 \bar{S_3}\bar{S_3} = -1$ } (-ve class)

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

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

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

After solving these equations we get –

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

$\alpha_1 = 5/2 = 2.5$ $\quad$ $\alpha_2 = -1,$ $\quad$ $\alpha_3 = -1$

Finding weighted vector $\bar{w} = \sum \alpha_i \bar{s}_i$

$$w = 2.5\begin{pmatrix}1\\0\\0\\1\end{pmatrix} - 1\begin{pmatrix}1\\-\sqrt{2}\\1\\1\end{pmatrix} - 1\begin{pmatrix}1\\\sqrt{2}\\1\\1\end{pmatrix}$$

$$= \begin{pmatrix}0.5\\0\\-2\\1/2\end{pmatrix}$$

From this, we need to remove bias so

$$w = \begin{pmatrix}0.5\\0\\-2\end{pmatrix} \quad \& \quad b = 0.5.$$

Eq^n of hyperplane, $y = wx + b$

$$= \begin{pmatrix}0.5\\0\\-2\end{pmatrix} x + 0.5$$

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

Given: $\hat{\omega} = [\omega_1, \omega_2, \omega_3]^T$.

For inequality constraint we apply KKT. Here we have 3 vectors in 3D space & all are support vectors so equality holds & we can apply Lagrange.

3 - constraints so 3 parameters. $(\lambda_1 \lambda_2 \lambda_3)$

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

$$\nabla_w L = \frac{\partial L}{\partial w} = w + \sum_{i=1}^{3} \lambda_i y_i \phi(x_i) = 0$$

$$\nabla_b L = \frac{\partial L}{\partial b} = \sum_{i=1}^{3} \lambda_i y_i = 0.$$

Putting $\phi(x_i)$ from part (b) we get,

$$w_1 + \lambda_1 - \lambda_2 - \lambda_3 = 0 \quad - (1)$$

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

$$w_3 - \lambda_2 - \lambda_3 = 0 \quad - (3)$$

$$\lambda_1 - \lambda_2 - \lambda_3 = 0 \quad - (4)$$

From above eq^n (1), (4) we get $\boxed{w_1 = 0}$

now, $\boxed{b = 1}$

$$-\sqrt{2}w_2 + w_3 + b = -1 \quad - (4)$$

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

From eq (4) & (5); we get $\boxed{w_2 = 0}$ $\boxed{w_3 = -2}$

Now the weights are $(0, 0, -2)^T$ & $b = 1$.

& margin $= \frac{2}{\|w\|_2^2} = \frac{2}{4} = \frac{1}{2}$

Scanned with CamScanner

d) New constraint = $y_i(w^T\phi(x_i)+b) \geq p$, $i=1,2,3$, $p \geq 1$
The value of b & w will change due to new constraint, rest all the part will remain same as in part (c) $\boxed{b=p}$ & $w=(0,0,-2p)^T$.

So we'll have the same classifier in both the cases. Only the eq^n of hyperplane is scaled by a factor of p & it's the property of sep. hyperplane eq^n that it is scale invariant, means the eq^n of hyperplane doesn't change by scaling the eq^n.

e) Yes it is true for any dataset as it follows scale invariance property. For constraints in (d) we can define new wt vectors $\bar{w}=w/p$ & $\bar{b}=b/p$. so the constraints in new variables becomes $(y_i(\bar{w}^T\phi(x_i)+\bar{b})=1$

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

$$s/t \quad y_i(\bar{w}^T\phi(x_i)+\bar{b}) \geq 1, \; i=1,2,3$$

Because p^2 is constant multiplying the fun^c $\|\bar{w}\|_2^2$ it doesn't change the optimal value. As $w^Tx+b \geq 0 = p\bar{w}^T + p\bar{b} \geq 0 \Rightarrow$ both gives same hyperplane & describe same classifier.

Ques 4) a) $K(x,x') = cK'(x,x')$, $c>0$
Let feature map of k' be $K'(x,x') = \phi(x)^T\phi(x')$ — ①
where $\phi'(x) = (\phi_1'(x) \ldots \phi_{N_T}'(x))$
From this we can write,

$$K(x,x') = \sqrt{c}\,\phi(x)^T \cdot \sqrt{c}\,\phi(x')$$
$$= c\,\phi(x)^T \cdot \phi(x')$$

$\boxed{K(x,x') = c\,K_1(x,x')}$ as given K_1 = valid kernal so this is also a valid kernal.

Scanned with CamScanner

___/___/___

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

LHS: $K(x,x') = z^T \cdot K(x,x') \cdot z \geq 0, \forall z \in R^n$

from RHS, put value of $K(x,x')$

$= z^T [k^1(x,x') + k^2(x,x')] z$

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

Given k^1 & k^2 = valid kernals & their add^n is also greater than eq to zero i.e. (semi +ve definite)

Hence $K(x,x')$ is a valid kernal.

c) $K(x,x') = f(x)\, k'(x,x') \cdot f(x')$ where f is a func from $R^m \to R$

We know that $K'(x,x') = \phi'(x) \cdot \phi'(x')$

The given expression can be written as,

$f(x) \cdot \phi'(x) \cdot \phi'(x') \cdot f(x')$ — ①

Lets assume, $\phi^2(x) = f(x) \cdot \phi'(x)$

$\phi^2(x') = f(x')\, \phi'(x')$

The eq^n ① becomes $K^2(x,x') = \phi^2(x)$

This is a valid kernal as stated in question.

Similarly $K(x,x')$ is a valid kernal.

d) $K(x,x') = k'(x,x')\, k^2(x,x')$

We can write $K'(x,x') = \phi'(x)^T \cdot \phi'(x')$

$K^2(x,x') = \phi^2(x)^T \cdot \phi^2(x')$

where, $\phi'(x) = (\phi_1'(x), \phi_2'(x) \dots \phi_n'(x))$

$\phi^2(x) = (\phi_1^2(x), \phi_2^2(x) \dots \phi_m^2(x))$

From this exp, we define

$\phi^3(x) = [\phi_1^1(x)\ \phi_1^2(x) \dots \phi_1^1(x)\ \phi_m^2(x)\ \phi_2^1(x)\phi_1^2(x) \dots \phi_m^2(x)]$

$\Rightarrow K(x,x') = \phi^3(x)^T \cdot \phi^3(x')$

As $\phi^3(x)$ made from $\phi'(x)$ & $\phi^2(x)$ which are valid kernals. feature vectors so we can say $\phi^3(x)$ is a feature vec of valid kernal $K(x,x')$

$K(x,x')$ is a valid kernal.