Machine learning
Assignment - 3

MOHIT GUPTA
MT22112

Que 3.

(a) Maximize $f(x,y) = xy$.

S.t. $x + y^2 \leq 2$

$x, y > 0$

Sol^n: To prove a function to be maxi., we must prove it to be concave.

But $f(x,y)$ is neither convex nor concave. which we can see from the hessian matrix.

$H_f = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, having two eigen values.

$\lambda_1 = 1, \lambda_2 = -1$.

for concavity / convexity H must be +ve semi-definite having all eigenvalues ≥ 0.

But the feasible region is bounded, so a global maximum must exist.

→ we can write the given constraint as,

$g(x,y) = x^2 + y^2 - 2 \leq 0$

The necessary conditions for KKT is,

$\nabla f(x,y) + \lambda \nabla g(x,y) = 0$ —— ①

$\lambda g(x,y) = 0$ —— ②

$g(x,y) \leq 0$ —— ③

$\lambda \geq 0$ —— ④

Here, we have only one constraint so, we have only one λ.

Calculating the constraints,

$$\left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right] + \lambda \left[\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}\right] = 0$$

② $[-y, -x] + \lambda[1, 2y] = 0$ (considering $f(x,y) = -xy$)

$$-y + \lambda = 0 \quad \text{——— } 1.a$$
$$-x + 2\lambda y = 0 \quad \text{——— } 1.b$$
$$\lambda(x + y^2 - 2) = 0 \quad \text{——— } 2.a$$
$$x + y^2 - 2 \leq 0 \quad \text{——— } 3.a$$
$$\lambda \geq 0 \quad \text{——— } 4.a$$

So, here we have two choices for λ to find maximum values for x & y.

<u>case 1</u> : $\lambda = 0$

from <u>eq 1.a</u>, & <u>eqn 1.b</u>

we get $\boxed{x = 0}$ & $\boxed{y = 0}$

<u>case 2</u> : $\lambda \neq 0$

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

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

put x & y in eqn 3.a

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

but from the 4.a, λ cannot be $-$ve

$$\lambda = \sqrt{\frac{2}{3}}$$

from this, $x = 2\left(\sqrt{\frac{2}{3}}\right)^2 = \frac{4}{3}$, $y = \sqrt{\frac{2}{3}}$

So from this we have two points, $(0,0)$, $\left(\frac{4}{3}, \sqrt{\frac{2}{3}}\right)$

from this, the global maximum is at $\left(\frac{4}{3}, \sqrt{\frac{2}{3}}\right)$.

as $x, y > 0$

& maximum value of $f(x,y)$ with given constraint is,

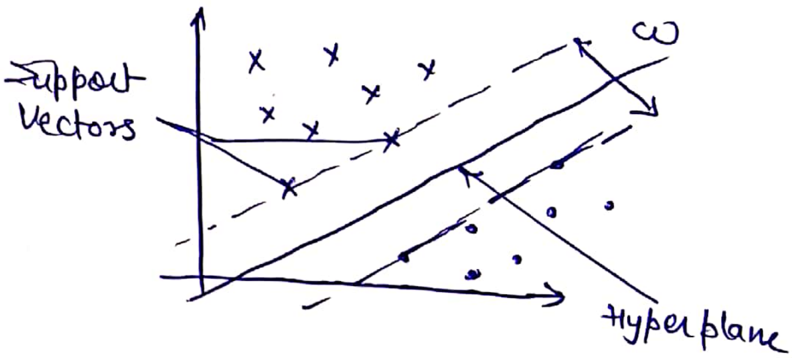
$$f\left(\frac{4}{3}, \sqrt{\frac{2}{3}}\right) = \frac{4}{3} \cdot \sqrt{\frac{2}{3}}.$$

b) True,

for any given training dataset, the model/hyperplane produced by SVM is always greater than or equal to any model that perfectly classifies the data, because SVM is called the maximum margin classifier, the main objective of SVM is to find the optimal hyper-plane with maximum margin from the closest support vectors

All the points on the hyperplane must satisfy this equation,

$$w^T x = 0$$



The aim of SVM (maxi. margin classifier) is to maximize $\frac{2}{\|w\|}$.

That is how, SVM finds out the ~~maxim~~ may which is greater than or equal to any other hyperplane produced to classify the same dataset.

4

a) $K(x,x') = ck'(x,x')$, $c > 0$

Soln Let feature map of k' be

$$k'(x,x') = \phi(x)^T . \phi(x') \quad \text{——} \quad (i)$$

where $\phi'(x) = (\phi_1'(x), \dots \phi_{N_1}'(x))$

from this we can write,

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

$$= c\,\phi(x)^T . \phi(x')$$

$$\boxed{K(x,x') = c\,K_1(x,x')} \quad \{\text{from } 1)$$

As it is given that k_1 is a valid kernel, then $k(x,x')$ is also a valid kernel.

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

Soln : LHS : $K(x,x')$

$$= \quad z^T . K(x,x') . z \geq 0 \quad , \forall z \in R^n$$

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

$$= z^T \left[k'(x,x') + k^2(x,x') \right] z$$

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

Given k^1 & k^2 are valid kernals, & their addition is also greater than equal to zero. i.e. (+ve semi definite)

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

(c) $K(x, x') = f(x) K'(x, x') . f(x')$

where f is a function from $R^m \rightarrow R$

Sol^n : We can write,

$$K'(x, x') = \phi'(x) . \phi'(x')$$

The given expression can be written as,

$$= f(x) . \phi'(x) . \phi'(x') . f(x') \longrightarrow ①$$

Let's assume.

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

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

The eqn ① becomes.

$$K^2(x, x') = \phi^2(x) . \phi^2(x')$$

This is a valid kernel, as given in the question, similarly we can say that $k(x, x')$ is a valid kernel.

(d) $K(x, x') = K'(x, x') . K^2(x, x')$

Sol^n We can write,

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

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

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, we can write it as,

$$[\phi_1'(x) \phi_1^2(x) . \phi_1'(x) \phi_2^2(x) - \dots \phi_1'(x) \phi_m^2(x) \dots]$$

from the above expansion, we define

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

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

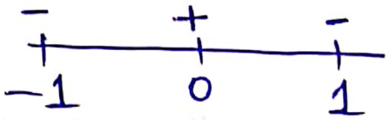
As $\phi^3(x)$ made from $\phi^1(x)$ & $\phi^2(x)$ which are valid kernels feature vectors, so we can say that $\phi^3(x)$ is a feature vector of a valid kernel $K(x, x')$

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

Que 5

class	x
+	0
−	−1
−	+1

(a) plotting the point on a 1D line



from the visualization as well, we cannot draw a linearly seperable hyperplane which can separate the two classes $\{+, -\}$.

So, the two classes are not linearly separable.

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

from the above values of x

$\phi(0) = (1, 0, 0)$

$\phi(-1) = (1, -\sqrt{2}, 1)$

$\phi(1) = (1, \sqrt{2}, 1)$

Now, these three points are separable in 3-dimension.

Now finding the separating hyperplane.

let us assume, $\phi(0)$, $\phi(-1)$ & $\phi(1)$ are the support vector & add a bias of 1, The support vector becomes

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

$$\alpha_1 S_1 S_1 + \alpha_2 S_1 S_2 + \alpha_3 S_1 S_3 = 1 \quad] \text{ for class } (+)$$

$$\left.\begin{aligned} \alpha_1 S_1 S_2 + \alpha_2 S_2 S_2 + \alpha_3 S_2 S_3 &= -1 \\ \alpha_1 S_1 S_3 + \alpha_2 S_2 S_3 + \alpha_3 S_3 S_3 &= -1 \end{aligned}\right] \text{ for class } (-)$$

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

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

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

After solving this, we get.

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

$$\alpha = \frac{5}{2}, \quad \alpha_2 = -1, \quad \alpha_3 = -1$$

Now finding the weight vector

$$w = \sum_{i=1}^{3} \alpha_i s_i$$

$$= \frac{5}{2}\begin{pmatrix}1\\0\\0\\1\end{pmatrix} + (-1)\begin{pmatrix}1\\-\sqrt{2}\\1\\1\end{pmatrix} + (-1)\begin{pmatrix}1\\\sqrt{2}\\1\\1\end{pmatrix}$$

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

from this, we get $w = \begin{pmatrix}1/2\\0\\-2\end{pmatrix}$, $b = 1/2$

eqn of hyperplane,

$$y = wx + b$$

$$y = \begin{pmatrix}1/2\\0\\-2\end{pmatrix}(1 \ \sqrt{2}x \ x^2) + \frac{1}{2}$$

(c) $\min_{w,b} \frac{1}{2}\|w\|_2^2$ s.t

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

Given : $w = (w_1, w_2, w_3)^T$

Here we have 3 constraints, & should have 3 lagrange multipliers $(\lambda_1, \lambda_2, \lambda_3)$.

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

differentiate & equate to zero,

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

$$\frac{\partial \mathcal{L}(w,\lambda)}{\partial b} = \sum_{i=1}^{3} \lambda_i y_i = 0$$

now putting $\phi(x_i)$ from above part (b), we get

$$w_1 + \lambda_1 - \lambda_2 - \lambda_3 = 0 \quad \text{——— } a$$

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

$$w_3 - \lambda_2 - \lambda_3 = 0 \quad \text{——— } c$$

$$\lambda_1 - \lambda_2 - \lambda_3 = 0 \quad \text{——— } d$$

from above equations. a & d, we get $\boxed{w_1 = 0}$

now, $\boxed{b = 1}$ ——— (e)

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

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

from eq^n (f) & (g), we get

$$\boxed{w_2 = 0}, \boxed{w_3 = -2},$$

from this, we can say that the weights are $(0,0,-2)^T$

& $b = 1$.

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

(d). constraint is

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

Changing the constraints only changes the value of b & $w = (0,0,-2\rho)^T$ from the above solution.

So, we have the same classifier in both the cases, only the equation of hyperplane is scaled by a factor ρ. But it is the property that separating hyperplane equation is scale invariant, means the hyperplane doesn't change by scaling the equation.

(e) Yes, it is True for any dataset because it follows the property of scale invariance. ~~Means~~

for the constraint mentioned in part (d),

$y_i(w^T\phi(x_i)+b) \geq \rho$, we can define new wt. vector $\bar{w} = w/\rho$ & $\bar{b} = b/\rho$.

Then the constraint with new variables

become, $$y_i(\bar{w}^T\phi(x_i)+\bar{b}) = 1.$$

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

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

Since ρ^2 is constant multiplying the func^n $\|\bar{w}\|_2^2$. it doesn't change the optimal value. As

$$w^Tx+b \geq 0 \equiv \rho\bar{w}^Tx+\rho\bar{b} \geq 0$$

Both gives the same hyperplane & describe the same classifier.