Tamoghna Chattopadhyay
EE 559.

HOMEWORK - 9

Q1.

(a) Yes, if the above set of constraints is satisfied, all the training data will be correctly classified. This is because SVM finds the closest points of two classes and draws the best hyperplane possible so that the margin from both is maximum. This generally tends to properly classify linearly separable data fully.

(b) $L(\underline{w}, w_0, \underline{\lambda})$

$$= \frac{1}{2} \|w\|^2 - \sum_{i=1}^{N} \lambda_i [z_i(\underline{w}^T \underline{u}_i + w_0) - 1]$$

$$\lambda_i \geq 0 \quad \forall i \qquad \text{—— (1)}$$

$$\lambda_i [z_i(\underline{w}^T \underline{u}_i + w_0) - 1] = 0 \quad \forall i$$

(c) • $\nabla_{\underline{w}} L = 0$

$$\Rightarrow \|w\| - \sum_{i=1}^{N} \lambda_i z_i \underline{u}_i = 0$$

$$\Rightarrow \underline{w}^* = \|w\| = \sum_{i=1}^{N} \lambda_i z_i \underline{u}_i \qquad \text{—— (2)}$$

• $\frac{\partial L}{\partial w_0} = 0$

$$\therefore -\sum_{i=1}^{N} \lambda_i z_i = 0$$

$$\Rightarrow \sum_{i=1}^{N} \lambda_i z_i = 0 \qquad \text{—— (3)}$$

Substituting (2) and (3) in (1),

$$L_D = \frac{1}{2} w \cdot w - \sum_{i=1}^{N} \lambda_i z_i w u_i - w_0 \sum_{i=1}^{N} \lambda_i z_i + \sum_{i=1}^{N} \lambda_i$$

$$= \frac{1}{2} \sum_{i=1}^{N} \lambda_i z_i w u_i - \sum_{i=1}^{N} \lambda_i z_i w u_i + \sum_{i=1}^{N} \lambda_i$$

$$= -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j u_i^T u_j z_i z_j + \sum_{i=1}^{N} \lambda_i$$

with $\sum_{i=1}^{N} \lambda_i z_i = 0$

$\lambda_i \geq 0 \quad \forall i$

$\underline{w}^* = \sum_{i=1}^{N} \lambda_i z_i \underline{u}_i$

$\lambda_i \left[z_i \left(\underline{w}^{*T} \underline{u}_i + w_0 \right) - 1 \right] = 0 \quad \forall i$

82. $U_1 = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \in S_1$ $\qquad U_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \in S_2$

$z U_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

$U_1^T U_1 = 1 \qquad U_2^T U_2 = 1 \qquad U_1^T U_2 = U_2^T U_1 = 0$

a) $L_D(\underline{\lambda}, \mu) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\left[\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j z_i z_j u_i^T u_j\right] + \mu\left(\sum_{i=1}^{N} z_i \lambda_i\right)$

$= \lambda_1 + \lambda_2 + \mu(z_1\lambda_1 + z_2\lambda_2)$

$- \frac{1}{2}\left\{ \lambda_1^2 z_1^2 u_1^T u_1 + \lambda_1\lambda_2 z_1 z_2 U_1^T U_2 + \lambda_2\lambda_1 z_2 z_1 U_2^T U_1 + \lambda_2^2 z_2^2 U_2^T U_2 \right\}$

$= \lambda_1 + \lambda_2 - \frac{1}{2}\left\{\lambda_1^2 + \lambda_2^2\right\} + \mu\lambda_1 - \mu\lambda_2$

$= \lambda_1(1+\mu) + \lambda_2(1-\mu) - \frac{1}{2}(\lambda_1^2 + \lambda_2^2)$

So,

$\frac{\partial L_D}{\partial \lambda_1} = 1 + \mu - \lambda_1 = 0 \Rightarrow \lambda_1 = 1 + \mu$

$\frac{\partial L_D}{\partial \lambda_2} = 1 - \mu - \lambda_2 = 0 \Rightarrow \lambda_2 = 1 - \mu$

$\frac{\partial L_D}{\partial \mu} = \lambda_1 - \lambda_2 = 0 \Rightarrow \lambda_1 = \lambda_2$

so, $1 + \mu = 1 - \mu \Rightarrow \mu = 0$

$\Rightarrow \lambda_1 = \lambda_2 = 1$

So,

$$\underline{w}^* = \lambda_1 z_1 \underline{v}_1 + \lambda_2 z_2 \underline{v}_2$$

$$= 1 \times 1 \times \begin{bmatrix} -1 \\ 0 \end{bmatrix} + 1 \times (-1) \times \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

For w_0,

$$\lambda_1 \left[z_1 \left([-1 \; -1] \begin{bmatrix} -1 \\ 0 \end{bmatrix} + w_0 \right) - 1 \right] = 0$$

$$\therefore \; 1 \times 1 \times 1 + w_0 - 1 = 0$$

$$\therefore \; w_0 = 0$$

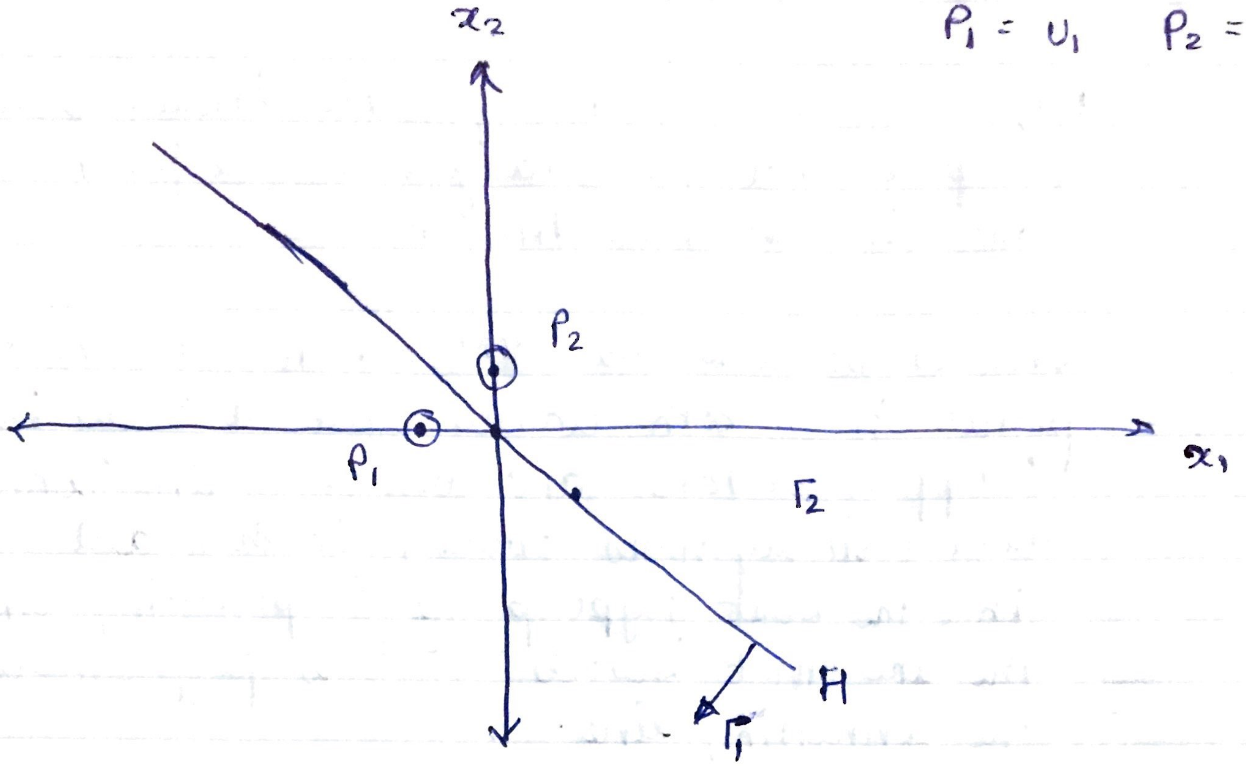
Here, $\underline{\lambda} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ satisfies the restrictions from Problem 1 (c) (ii).

$$\because \; \underline{\lambda} > 0$$

$$\sum_{i=1}^{2} \lambda_i z_i = 0$$

PTO.

$P_1 = U_1 \qquad P_2 = U_2$



$g(x) = W_0 + W_1 x_1 + W_2 x_2 = 0$

$\therefore \quad 0 - x_1 - x_2 = 0$

$\therefore \quad x_1 + x_2 = 0 \quad \longleftarrow g(x)$

Decision Boundary H.

b) Distance of P_1 from line

$$= \frac{|-1+0|}{\sqrt{1^2+1^2}} = \frac{1}{\sqrt{2}}$$

Distance of P_2 from line

$$= \frac{|0+1|}{\sqrt{1^2+1^2}} = \frac{1}{\sqrt{2}}$$

so, the line is equidistant from both points P_1 and P_2.

No, there is no other possible linear boundary in u-space that gives larger values for both margin distances than H.

This is because the SVM searches for the closest points from both classes which it calls the 'support vectors' and draws a line connecting them (in layman's terms). It then declares that the best hyperplane (separating line) is the line that bisects - and is perpendicular to - the connecting line.

Here, this condition is fulfilled by H which is the perpendicular bisector of line joining the two points - hence best possible hyperplane with highest margin.

Q3. $J(\underline{w}) = \frac{1}{2}\|\underline{w}\|^2 + C\sum_{i=1}^{N} \xi_i$

s.t. $z_i(\underline{w}^T u_i + w_0) \geq 1 - \xi_i \quad \forall i$

and, $\xi_i \geq 0 \quad \forall i$

a) $L(x, \lambda) = f(x) - \lambda h(x)$ s.t. $h(x) \geq 0$

So, here,

$$L(\underline{w}, w_0, \xi, \underline{\lambda}, \underline{\mu})$$

$$= \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\lambda_i\left[z_i(\underline{w}^T\underline{u}_i + w_0) - 1 + \xi_i\right] - \sum_{i=1}^{N}\mu_i\xi_i \quad \text{—— (1)}$$

s.t.

$\lambda_i \geq 0 \quad \forall i$

$\mu_i \geq 0 \quad \forall i$

$\mu_i \xi_i = 0 \quad \forall i$

$\lambda_i\left[z_i(\underline{w}^T\underline{u}_i + w_0) - 1 + \xi_i\right] = 0 \quad \forall i$

b) • $\nabla_{\underline{w}} L = 0$

$\therefore \|w\| - \sum_{i=1}^{N}\lambda_i z_i \underline{u}_i = 0$

$\therefore \underline{w}^* = \|w\| = \sum_{i=1}^{N}\lambda_i z_i \underline{u}_i \quad \text{—— (2)}$

• $\frac{\partial L}{\partial w_0} = 0 \Rightarrow \sum_{i=1}^{N}\lambda_i z_i = 0 \quad \text{—— (3)}$

• $\nabla_{\xi} L = 0$

$\therefore \quad C - \underline{\lambda} - \underline{\mu} = 0$

$\therefore \quad \underline{\lambda} + \underline{\mu} = C$

From here, if $\lambda < C \longrightarrow \xi_i = 0$
and $\xi_i \geq 0$ — if $\lambda = C \Rightarrow \mu = 0$

So, $0 \leq \lambda_i \leq C \quad \forall i$ — ④

Substituting ②, ③, ④ in ①,

$$L_D(\underline{\lambda}) = \frac{1}{2}\left(\sum_{i=1}^{N} \lambda_i z_i \underline{u}_i\right)\left(\sum_{j=1}^{N} \lambda_j z_j \underline{u}_j\right) + C\sum_{i=1}^{N} \xi_i$$

$$- \sum_{i=1}^{N} \lambda_i z_i \left(\sum_{j=1}^{N} \lambda_j z_j \underline{u}_j\right)^T u_i$$

$$- 0 + \sum_{i=1}^{N} \lambda_i - \sum_{i=1}^{N} (\lambda_i + \mu_i)\xi_i$$

$$= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j z_i z_j \underline{u}_i^T \underline{u}_j + \sum_{i=1}^{N} \lambda_i$$

$$- C\sum_{i=1}^{N} \xi_i + C\sum_{i=1}^{N} \xi_i$$

$$= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j z_i z_j \underline{u}_i^T \underline{u}_j + \sum_{i=1}^{N} \lambda_i$$

s.t. $\quad 0 \leq \lambda_i \leq C \quad \forall i \quad$ with $\sum_{i=1}^{N} \lambda_i z_i = 0$

$\lambda_i \geq 0 \quad \forall i$

$\mu_i \geq 0 \ \forall i \qquad \mu_i \xi_i = 0 \ \forall i$

$\lambda_i \left[z_i (\underline{w}^T \underline{u}_i + w_0) - 1 + \xi_i \right] = 0 \ \forall i$