Saurabh Tirodkar 3589406164

EE559 - HW #7

Q.1) Minimize $J(\underline{w}) = \frac{1}{2} \|w\|^2$

s.t. $z_i(\underline{w}^T \underline{u}_i + w_0) - 1 \geq 0 \quad \forall i$

Linearly separable in u space, non-augmented.

a) In general,

$$L(\underline{x}, \lambda) = \underbrace{f(x)}_{\text{Original func}} + \lambda \underbrace{g(\underline{x})}_{\text{Equality Constraints}}$$

When constraint is satisfied,

$$L(\underline{x}, \lambda) = f(x).$$

So, here, if the given set of constraints $[z_i(w^T u_i + w_0) - 1 \} \geq 0]$ is satisfied, all the training data will be correctly classified.

SVM finds the closest points & classify using best hyperplane possible.

As given the data is linearly separable in u-space, SVM correctly classify all the training data.

b) $L(\underline{w}, w_0, \underline{\lambda})$ for minimization = ?

λ_i : 0, 1, 2, . . . N.

KKT cond^n = ?

→

$$L(\underline{w}, w_0, \lambda) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \lambda_i \left[z_i (w^T u_i + w_0) - 1 \right]$$

— ①

KKT cond^n

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

2) $(z_i (w^T u_i + w_0) - 1) \geq 0$

3) $\lambda_i [z_i (w^T u_i + w_0) - 1] = 0 \quad \forall i$

c) Dual Lagrangian func^n L_D.

Steps 1. Minimize L wrt weights.

2. Use eq^n from ① & derive L_D.

③

$$\rightarrow \nabla_w L = \nabla_w \left[\frac{1}{2} \|w\|^2 - \sum \lambda \left(z_i (w^T u_i + w_0) - 1 \right) \right]$$

$$= \cancel{\frac{1}{2}} \cdot \cancel{2} \|w\| - \sum_{i=1}^{N} \lambda_i z_i u_i = 0$$

$$w^* = \|w\| = \sum_{i=1}^{N} \lambda_i z_i u_i \quad — ②$$

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

$$\therefore \quad 0 - \sum_{i=1}^{N} \lambda_i (0 + z_i)$$

$$\Rightarrow \sum_{i=1}^{N} \lambda_i z_i = 0 \quad — ③$$

Sub^n ② & ③ in ①.

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

$$- \underbrace{\sum_{i=1}^{N} \lambda_i z_i w_0}_{0} + \sum_{i=1}^{N} \lambda_i$$

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

P.T.O

KKT cond^n

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

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

$$w^* = \sum_{i=1}^{N} \lambda_i z_i u_i$$

$$(z_i (w^T u_i + w_0) - 1) \geq 0$$

$$L_D(\lambda) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\lambda_i \lambda_j z_i z_j u_i^T u_j)$$

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

$$+ \sum_{i=1}^{N} \lambda_i$$

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

Q.2) SVM learning

$u_1 = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \in S_1 \qquad u_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \in S_2.$

Using eeflection,

$$z u_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

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

2 class, 2 data pts.

$$= \lambda_1 + \lambda_2 + \mu(z_1\lambda_1 + z_2\lambda_2) - \frac{1}{2}\Big(\lambda_1^2 z_1^2 u_1^T u_1 + \underbrace{\lambda_1\lambda_2 z_1 z_2 u_1^T u_2}_{0} + \underbrace{\lambda_2\lambda_1 z_2 z_1 u_2^T u_1}_{0} + \lambda_2^2 z_2^2 u_2^T u_2\Big)$$

$$= \lambda_1 + \lambda_2 + \mu z_1 \lambda_1 + \mu z_2 \lambda_2 - \frac{1}{2}\left(\lambda_1^2 z_1^2 + \lambda_2^2 z_2^2\right)$$

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

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

Finding λ_1 & λ_2.

$$\frac{\partial L_D}{\partial \lambda_1} = 1+\mu - \frac{1}{2} \cdot 2\lambda_1 = 0$$

$$\Rightarrow \lambda_1 = 1+\mu.$$

$$\frac{\partial L_D}{\partial \lambda_2} = 1-\mu - \frac{1}{2} \cdot 2\lambda_2$$

$$\Rightarrow \lambda_2 = 1-\mu$$

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

$$\therefore 1+\mu = 1-\mu$$

$$\Rightarrow \mu = 0 \quad \therefore \lambda_1 = \lambda_2 = 1$$

$$w^* = \sum_{i=1}^{N} \lambda_i z_i u_i \quad \text{from Q. 1.}$$

Here,

$$w^* = \lambda_1 z_1 u_1 + \lambda_2 z_2 u_2$$

$$= (1).(1) \begin{pmatrix} -1 \\ 0 \end{pmatrix} + 1(-1) \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

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

$$\boxed{w^* = \begin{pmatrix} -1 \\ -1 \end{pmatrix}}$$

$w_0 = ?$

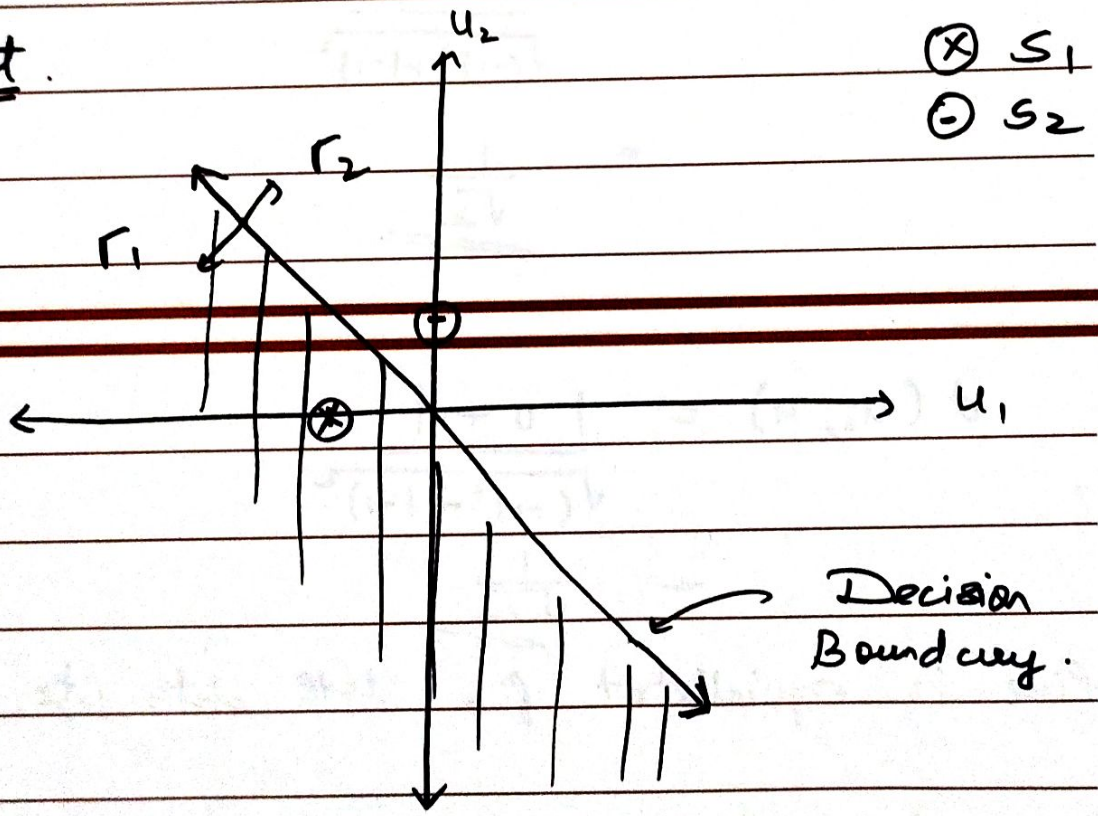
$$\lambda_1 (z_1 (w^T u_1 + w_0) - 1) = 0$$

$$1 \cdot \left(1 \, (-1 \;\; -1) \cdot \begin{pmatrix} -1 \\ 0 \end{pmatrix} + w_0 - 1 \right) = 0$$

$$1 + w_0 - 1 = 0$$

$$\boxed{w_0 = 0}$$

Plot.



$$g(u) = w_1 u_1 + w_2 u_2 + \underset{\substack{\parallel \\ 0}}{w_0} = 0$$

$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$ $\therefore$ $-u_1 - u_2 = 0$

& $w_0 = 0$

$u_1 + u_2 = 0$ ← Dec-Boundary

b) $H = u_1 + u_2 = 0$.

$d(u_1, H) = ?$

$d(u_2, H) = ?$

$$d(u_0, H) = \frac{g(u_0)}{\|w\|}$$

So, $d(u_1, H) = \dfrac{|-1+0|}{\sqrt{(-1)^2+(-1)^2}}$

$$= \frac{1}{\sqrt{2}}$$

$d(u_2, H) = \dfrac{|0+1|}{\sqrt{(-1)^2+(-1)^2}}$

$$= \frac{1}{\sqrt{2}}$$

Line is equidistant from both data pts.

Other Possibility Comment

No, other possible linear boundary in u-space that would give larger values for both distances than H.

Since, SVM gives the most optimal solution & the decision boundary is plotted such that the data points are ~~equ~~ classified in best possible hypeaplane.