Sourabh Tirodkar EE 559

HW # 5

Q.2) $\{0, 1, -1, 2\} \in S_1$

$\{(1,1,1,1), (2,1,1,1)\} \in S_2$

$\{(-1,1,0,-1)\} \in S_3$

Using augmented space,

given weights $w^{(1)}(0) = -1$

$w^{(2)}(0) = 1$

$w^{(3)}(0) = 0$

$\therefore \; w_1(0) = (-1,-1,-1,-1,-1)$

$w_2(0) = (1,1,1,1,1)$

$w_3(0) = (0,0,0,0,0)$

Decision Rule:

$x \in S_1$ iff $g_1(x) > g_2(x)$ & $g_1(x) > g_3(x)$

$x \in S_2$ iff $g_2(x) > g_1(x)$ & $g_2(x) > g_3(x)$

$x \in S_3$ iff $g_3(x) > g_1(x)$ & $g_3(x) > g_2(x)$

$$g_k(x) = w_k^T x$$

$x_1 = (1, 0, 1, -1, 2)$

$g_1(x) = -3$

$g_2(x) = 3$

$g_3(x) = 0$

$w_1(1) = w_1(0) + x_1$

$= (0, -1, 0, -2, 1)$

$w_2(1) = w_2(0) - x_1$

$= (0, 1, 0, 2, -1)$

$w_3(1) = w_3(0)$

$= (0, 0, 0, 0, 0)$

$x_2 = (1, 1, 1, 1, 1)$

$g_1(x) = -2$

$g_2(x) = 2$

$g_3(x) = 0$

$w_1(2) = w(1)$

$= (0, -1, 0, -2, 1)$

$w_2(2) = w_2(1)$

$= (0, 1, 0, 2, -1)$

$w_3(2) = (0, 0, 0, 0, 0)$

$x_3 = (1, 2, 1, 1, 1)$

$g_1(x) = -3$

$g_2(x) = 3$

$g_3(x) = 0$

$w_1(3) = w_1(2)$

$= (0, -1, 0, -2, 1)$

$w_2(3) = w_2(2)$

$= (0, 1, 0, 2, -1)$

$w_3(3) = w_3(2) = (0, 0, 0, 0, 0)$

$x_4 = (1, -1, 1, 0, -1)$

$g_1(x) = 0$

$g_2(x) = 0$

$g_3(x) = 0$

$w_1(4) = w_1(3) + x_4$

$= (1, -1, 1, 0, -1)$

$w_2(4) = w_2(3) - x_4$

$= (-1, 2, -1, 2, 0)$

$w_3(4) = (0, -1, 0, -2, 1)$

Discriminant func^n

$$\boxed{\begin{aligned} g_1(x) &= -x_1 - 2x_2 + x_4 \\ g_2(x) &= -1 + 2x_1 - x_2 + 2x_3 \\ g_3(x) &= 1 - x_1 + x_2 - x_4 \end{aligned}}$$

b) $\underline{x} = (1, x_1, x_2, 0, 0)$

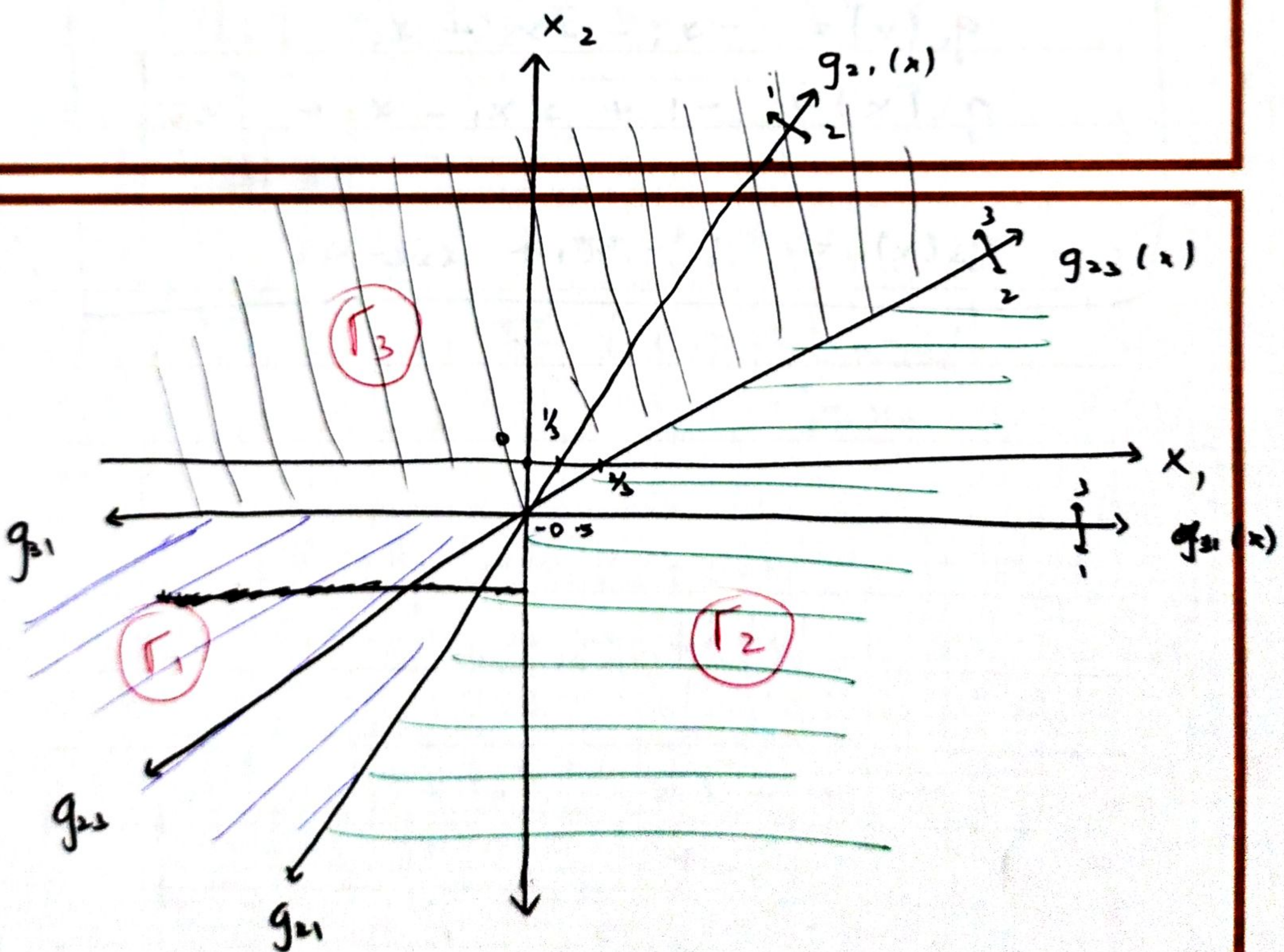
$g_1(x) = -x_1 \qquad g_2(x) = -1 + 2x_1 - x_2$

$g_3(x) = 1 - x_1 + x_2$

$g_{21}(x) = 3x_1 - x_2 - 1$

$g_{23}(x) = -2 + 3x_1 - 2x_2$

$g_{31}(x) = 1 + x_2$



Q.3) Widrow Hoff learning
– Basic Sequential G.D.

a) $J_n(w) = ?$

$\nabla_w J_n(\underline{w}) = ?$

$$J(\underline{w}) = \sum_{n=1}^{N} (w^T z_n x_n - b_n)^2$$

$$\boxed{J_n(\underline{w}) = (w^T z_n x_n - b_n)^2}$$

$$\therefore \nabla_w J_n(\underline{w}) = \nabla_w (w^T z_n x_n - b_n)^2$$

$$= \nabla_w \left[(w^T z_n x_n - b_n)^T (w^T z_n x_n - b_n) \right]$$

$$= \nabla_w \left[(z_n x_n)^T w w^T (z_n x_n) - w^T z_n x_n b_n^T - (z_n x_n)^T \underline{w}\, b_n + b_n^T b_n \right]$$

$$= \nabla_w \left[w^T (z_n x_n)(z_n x_n)^T w - 2 (z_n x_n)^T w\, b_n + b_n^T b_n \right]$$

$$= \left[2 (z_n x_n)(z_n x_n)^T w - 2 (z_n x_n)^T b_n \right]$$

$$= \left[2 (z_n x_n)(z_n x_n)^T w - 2 b_n (z_n x_n) \right]$$

$$= 2 z_n x_n \left(w^T (z_n x_n) - b_n \right)$$

$$\boxed{\therefore \nabla_w J_n(w) = 2 (z_n x_n) \left[w^T z_n x_n - b_n \right]}$$

b) For Seq. G.D.
Weight updation

$$w(i+1) = w(i) - \eta(i) \nabla J_n(\underline{w})$$

$$= w(i) - \eta(i) \cdot 2 (z_n x_n) \left[w^T z_n x_n - b_n \right]$$

To converge, let

$$\eta'(i) = 2 \eta(i)$$

$$\therefore w(i+1) = w(i) - \eta'(i)\,[z_n x_n)\,[w^T z_n x_n - b_n]$$

Steps in Seq. G.D.

1) Random Shuffling of data point.
2) Initialization for $w(0)$.
3) Def epoch index m as

$$n = \{1, 2 \ldots N\}$$

$$i = (m-1)N + n - 1$$

$$w(i+1) = w(i) - \eta(i)\,\nabla_w J_n(w)$$

$$w(i+1) = w(i) - \eta'(i)\, z_n x_n\,(b_n - w(i)^T z_n x_n)$$

$$\eta(i) \geq 0.$$

Need to converge M.

Q.1) 2-class perceptron with margin algorithm.

Prove - Convergence for linearly separable data.

Assumptions:

$\eta(i) = \eta = \text{const} > 0 \quad \forall i$

Seq. G.D.

linearly separable.

Reflected data pts $z_n x_n \quad n = 1, 2, .. N.$

$\eta = 1$ w/o loss of generality.

Let $z_n x_n' = z_n \eta x_n \qquad \eta > 0.$

Algorithm:

$$\begin{cases} w(0) = \text{arbitrary.} \\ w(i+1) = w(i) + z_i x_i \, [\![w(i)^T z_i x_i < 0]\!] \end{cases}$$

$z_i x_i = \quad i = 0, 1, 2 ..$ cyclically ordered.

let $z_i x_i \rightarrow$ subset of training data that are mis classified.

<u>Algorithm</u> :

1) $\begin{cases} w(0) = \text{arbitary} \\ w(i+1) = w(i) + z^i x^i \end{cases}$

in which $w(i)^T z^i x_i \le \underline{b} \quad \forall i$ ← <u>for margin</u>.

if $\hat{w}$ is a sol^n, then $a\hat{w}$ is also a sol^n.

$$\hat{w}^T z_n x_n > b \quad \forall n$$

$$a\hat{w}^T z_n x_n - ab > 0 \quad \forall n \quad \underline{a > 0}.$$

For error measure,

$$\mathcal{E}_w(i) = \| w(i) - a\hat{w} \|_2^2 - b$$

$$w(i+1) - a(\hat{w}) = (w(i) - a\hat{w}) + z^i x^i$$

$$\| w(i+1) - a\hat{w} \|_2^2 = \| w(i) - a\hat{w} + z_i x_i \|_2^2$$

$$\| w(i+1) - a\hat{w} \|_2^2 - b = \| w(i) - a\hat{w} + z_i x_i \|_2^2 - b.$$

$$\|w(i+1) - a\hat{w}\|_2^2 - b$$
$$= \|w(i) - a\hat{w} + z_i^i x_i^i\|_2^2 - 2b + b.$$

$$= \|w(i) - a\hat{w}\|_2^2 + 2(w(i) - a\hat{w})^T z_i x_i + \|z_i x_i\|_2^2 - 2b + b$$

$$= \|w(i) - a\hat{w}\|_2^2 + 2(w(i)^T z_i^i x_i^i - b) - 2a\hat{w}^T z_i x_i + \|z_i x_i\|_2^2 + b$$

But, $\underbrace{w(i)^T z^i x^i}_{\text{misclassified pt.}} \leq 0$

$$\varepsilon w(i+1) \leq \|w(i) - a\hat{w}\|_2^2 - 2a(\hat{w} z_i^i x_i^i - b) + \|z^i x^i\|_2^2 + b - 2ab$$

But $2ab > 0$

$-2ab < 0$ drop.

$$\varepsilon w(i+1) \leq \|w(i) - a\hat{w}\|_2^2 - 2a(w z_i^i x^i - b) + \|z^i x^i\|_2^2 + b$$

let $z^i x^i = \|x^i\|_2$

$$\beta^2 = \max \|x^i\|_2^2$$

$$c = \min_k \{\hat{\omega}^T z_k x_k\} > 0$$

$$\varepsilon \omega(i+1) \leq \|\omega(i) - a\hat{\omega}\|_2^2 - 2ac + \beta^2 + b$$

Assume $a = \frac{b + \beta^2}{c}$

$$\varepsilon \omega(i+1) \leq \|\omega(i) - a\hat{\omega}\|_2^2 - 2b - \beta^2 + b$$

$$\leq \|\omega(i) - a\hat{\omega}\|_2^2 - \cancel{2b} - \beta^2 + \cancel{2b} - b$$

$$\leq \|\omega(i) - a\hat{\omega}\|_2^2 - b - \beta^2$$

Apply fencing argument:

$$b \leq \varepsilon\omega(i+1) \leq \varepsilon\,\omega(i) - \beta^2 \quad \forall i$$

For some i_0, we would have

$$\varepsilon\omega(i_0) < \beta^2$$

so that

$b < \varepsilon w(i_0+1) \leq \varepsilon w(i_0) - \beta^2 < b$

→ impossible.

⇒ Iteration must cease at $i = i_0 - 1$
(or sooner)

$\varepsilon w(i_0)$ is result of $(i_0 - 1)^{th}$ iteration.

⇒ Algorithm converges at a solⁿ weight vector at $(i_0 - 1)^{th}$ iteration.