EE559 HW Week6 1139093874 Pin-Hsuan Lee.

1. perceptron with margin algorithm $\Rightarrow$ $\underline{w}(i+1) = \underline{w}(i) + z^i \underline{x}^i \, [\![\underline{w}(i)^T z^i \underline{x}^i - b < 0]\!]$

use error measure $\mathcal{E}(i) = \| \underline{w}(i) - a\underline{w} \|_2^2$, a is adjustable

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

$\Rightarrow \| w(i+1) - a\underline{w} \|_2^2 = \| w(i) - a\underline{w} \|_2^2 + 2(\underline{w}(i) - a\underline{w})^T z^i \underline{x}^i + \| \underbrace{z^i}_{-1} \underline{x}^i \|^2$ $\quad (\underline{w}(i) z^i \underline{x}^i < 2b)$

$\leq \| w(i) - a\underline{w} \|_2^2 - 2a\underline{w}^T z^i x^i + 2b + \| x^i \|_2^2$

Let $c \triangleq \min_j \{ \hat{\underline{w}}^T z_j \underline{x}_j \} > 0$, $d^2 \triangleq \max_j \| \underline{x}_j \|_2^2$ = length of the longest data point.

$\Rightarrow \| \underline{w}(i+1) - a\underline{w} \|_2^2 \leq \| \underline{w}(i) - a\underline{w} \|_2^2 - 2ac + 2b + d^2$

choose $a = \dfrac{b + d^2}{c}$

$\Rightarrow 0 \leq \mathcal{E}(i+1) \leq \mathcal{E}(i) - d^2 \quad \forall i$

For some i_0, we would have $\mathcal{E}(i_0) < d^2$

so that $0 \leq \mathcal{E}(i_0+1) \leq \mathcal{E}(i_0) - d^2 < 0$ which is impossible.

Therefore, iteration must cease at $i = i_0 - 1$ (or sooner)

$\Rightarrow$ Algorithm converges at a solution weight vector at $(i_0-1)^{th}$ iteration or sooner.

2. (a) $\{(0,1,-1,2)\} \in S_1$, $\{(1,1,1,1),(2,1,1,1)\} \in S_2$, $\{(-1,1,0,-1)\} \in S_3$

$\underline{w}^{(1)}(0) = -1$, $w^{(2)}(0) = 1$, $w^{(3)}(0) = 0$,

$w^{(1)T}(0) \cdot [1,0,1,-1,2]^T = -3$ $\quad$ $w^{(2)T}(0) \cdot [1,0,1,-1,2]^T = 3$ Max $\quad$ $w^{(3)T}(0) \cdot [1,0,1,-1,2]^T = 0$

$w^{(1)}(1) = \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \\ -1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 1 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 0 \\ -2 \\ 1 \end{bmatrix}$ $\quad$ $w^{(2)}(1) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 1 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 2 \\ -1 \end{bmatrix}$ $\quad$ $w^{(3)}(1) = w^{(3)}(0) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

$w^{(1)T}(1) \cdot [1\ 1\ 1\ 1\ 1]^T = -2$ $\quad$ $w^{(2)T}(1) \cdot [1\ 1\ 1\ 1\ 1]^T = 2$ $\quad$ $w^{(3)T}(1) \cdot [1\ 1\ 1\ 1\ 1]^T = 0$

$w^{(1)}(2) = w^{(1)}(1)$ $\quad$ $w^{(2)}(2) = w^{(2)}(1)$ $\quad$ $w^{(3)}(2) = w^{(3)}(1)$

$w^{(1)T}(2) \cdot [1\ 2\ 1\ 1\ 1]^T = -3$ $\quad$ $w^{(2)T}(2) \cdot [1\ 2\ 1\ 1\ 1]^T = 3$ $\quad$ $w^{(3)T}(2) \cdot [1\ 2\ 1\ 1\ 1]^T = 0$

$w^{(1)}(3) = w^{(1)}(2)$ $\quad$ $w^{(2)}(3) = w^{(2)}(2)$ $\quad$ $w^{(3)}(3) = w^{(3)}(2)$

$w^{(1)T}(3) \cdot [1\ -1\ 1\ 0\ -1]^T = 0$ $\quad$ $w^{(2)T}(3) \cdot [1\ -1\ 1\ 0\ -1]^T = 0$ $\quad$ $w^{(3)T}(3) \cdot [1\ -1\ 1\ 0\ -1]^T = 0$

$w^{(1)}(4) = \begin{bmatrix} 0 \\ -1 \\ 0 \\ -2 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ -1 \\ -2 \\ 2 \end{bmatrix}$ $\quad$ $w^{(2)}(4) = w^{(2)}(3) = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 2 \\ -1 \end{bmatrix}$ $\quad$ $w^{(3)}(4) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ -1 \end{bmatrix}$

$\Rightarrow g^{(1)}(x) = -x_2 - 2x_3 + 2x_4 - 1$ $\quad$ $\Rightarrow g^{(2)}(x) = x_1 + 2x_3 - x_4$ $\quad$ $\Rightarrow g^{(3)}(x) = -x_1 + x_2 - x_4 + 1$

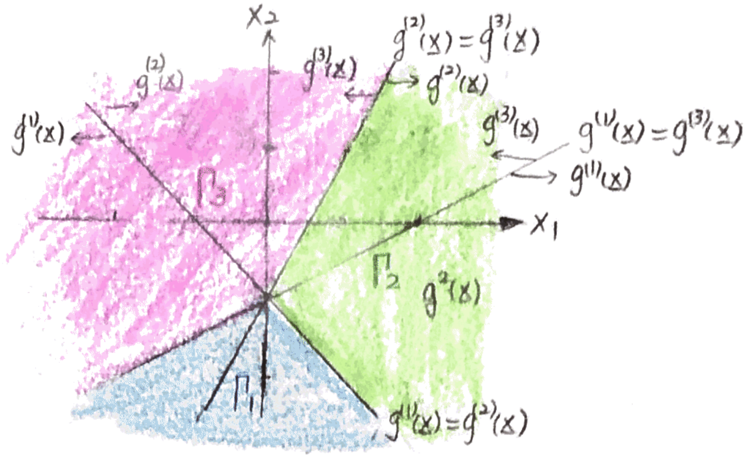
(b) $\underline{x} = (1, x_1, x_2, 0, 0)$

$g^{(1)}(\underline{x}) = -x_2 - 1$ $\quad$ $g^{(2)}(\underline{x}) = x_1$ $\quad$ $g^{(3)}(\underline{x}) = -x_1 + x_2 + 1$

$g^{(1)}(\underline{x}) = g^{(2)}(\underline{x}) \Rightarrow x_1 = -x_2 - 1 \Rightarrow x_1 + x_2 = -1$

$g^{(2)}(\underline{x}) = g^{(3)}(\underline{x}) \Rightarrow x_1 = -x_1 + x_2 + 1 \Rightarrow 2x_1 - x_2 = 1$

$g^{(3)}(\underline{x}) = g^{(1)}(\underline{x}) \Rightarrow -x_2 - 1 = -x_1 + x_2 + 1 \Rightarrow x_1 - 2x_2 = 2$



3.

(a)
$$J(\underline{w}) = \frac{1}{N}\| \underline{X}\underline{w} - \underline{b}\|_2^2 + \lambda \|\underline{w}\|_2^2$$
$$= \frac{1}{N}\left(\underline{w}^T\underline{X}^T\underline{X}\underline{w} - \underline{w}^T\underline{X}^T\underline{b} - \underline{b}^T\underline{X}\underline{w} + \underline{b}^T\underline{b}\right) + \lambda\, \underline{w}^T\underline{w}$$
$$\nabla_{\underline{w}} J(\underline{w}) = \frac{1}{N}\left(\underline{X}^T\underline{X}\underline{w} + \underline{X}^T\underline{X}\underline{w} - \underline{X}^T\underline{b} - \underline{X}^T\underline{b}\right) + \lambda(\underline{w} + \underline{w})$$
$$= \frac{2}{N}\left(\underline{X}^T\underline{X}\underline{w} - \underline{X}^T\underline{b}\right) + 2\lambda \underline{w}$$

(b)
$$\nabla_{\underline{w}} J(\underline{w}) = \frac{2}{N}\left(\underline{X}^T\underline{X}\underline{w} - \underline{X}^T\underline{b}\right) + 2\lambda \underline{w} = 0$$
$$\Rightarrow \left(\frac{2}{N}\underline{X}^T\underline{X} + 2\lambda \underline{I}\right)\underline{w} = \frac{2}{N}\underline{X}^T\underline{b}$$
$$\Rightarrow \left(\underline{X}^T\underline{X} + N\lambda \underline{I}\right)\underline{w} = \underline{X}^T\underline{b}$$
$$\Rightarrow \underline{w} = \left(\underline{X}^T\underline{X} + N\lambda\underline{I}\right)^{-1}\underline{X}^T\underline{b}$$

if $(\underline{X}^T\underline{X} + N\lambda\underline{I})$ is non-singular.

In Pseudoinverse solution
$$\underline{w} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{b}$$

By comparison, we can find that there's an additional term $N\lambda\underline{I}$ in the inverse matrix.

4. (a)
$$J(\underline{w}) = \sum_{n=1}^{N} J_n(\underline{w}) = \sum_{n=1}^{N}\left[\underline{w}^T z_n \underline{x}_n - \underline{b}_n\right]^2$$
$$J_n(\underline{w}) = \left[\underline{w}^T z_n \underline{x}_n - b_n\right]^2$$

(b)
$$\nabla_{\underline{w}} J_n(\underline{w}) = 2\left[\underline{w}^T z_n \underline{x}_n - \underline{b}_n\right] z_n \underline{x}_n \longrightarrow \text{"2" can be absorbed in } \eta(i).$$
$$w(i+1) = w(i) - \eta(i)\, \nabla_{\underline{w}} J_n(\underline{w}) \qquad \eta(i) \geq 0,\ w(0) = \text{arbitrary}$$
$$\Rightarrow w(i+1) = w(i) - \eta(i)\left[\underline{w}^T z_n \underline{x}_n - \underline{b}_n\right] z_n \underline{x}_n$$
$$\Rightarrow w(i+1) = w(i) + \eta(i)\left[\underline{b}_n - \underline{w}^T z_n \underline{x}_n\right] z_n \underline{x}_n$$

which is the same with the result of Widrow-Hoff learning algorithm.