jxu72364@usc.edu

1. Assumptions:

fixed increment $\eta(i) = \eta = \text{constant} > 0$

sequential gradient descent

Data points are linear separable

use reflected datapoints $z_n \underline{x}_n$, $n = 1, 2, \cdots N$

set $\eta = 1$

let $z^i \underline{x}^i$, $i = 0, 1, 2 \cdots$ be the subset of training datapoints that are misclassified at each iteration.

$$\Rightarrow \begin{cases} \underline{w}(0) = \text{arbitrary} \\ \underline{w}(i+1) = \underline{w}(i) + z^i \underline{x}^i \quad (i^{th} \text{ iteration}) \end{cases}$$

in which $\underline{w}(i)^T z^i \underline{x}^i \leq m \ \forall i$ $\quad m$ is the margin $\quad m > 0$

$\hat{\underline{w}}$ is a solution, then $a\hat{\underline{w}}$, $a > \frac{m}{\underline{w}(i)^T z^i \underline{x}^i}$ is also a solution:

$$\hat{\underline{w}}^T z_n \underline{x}^n > m \ \forall n$$

$$\Rightarrow a\hat{\underline{w}}^T z_n \underline{x}^n > m \ \forall n, \text{ if } a > \frac{m}{\underline{w}(i)^T z^i \underline{x}^i}$$

need "error measure" on $\underline{w}(i)$:

$$\varepsilon_{\underline{w}}(i) = \| \underline{w}(i) - a\hat{\underline{w}} \|_2^2$$

$$\underline{w}(i+1) - a\hat{\underline{w}} = (\underline{w}(i) - a\hat{\underline{w}}) + z^i \underline{x}^i \qquad a > \frac{m}{\underline{w}(i) z^i \underline{x}^i}$$

$$\| \underline{w}(i+1) - a\hat{\underline{w}} \|_2^2 = \| \underline{w}(i) - a\hat{\underline{w}} \|_2^2 + 2[\underline{w}(i) - a\hat{\underline{w}}]^T z^i \underline{x}^i + \| z^i \underline{x}^i \|_2^2$$

$$\underline{w}(i)^T z^i \underline{x}^i \leq m$$

$$\| \underline{w}(i+1) - a\hat{\underline{w}} \|_2^2 = \| \underline{w}(i) - a\hat{\underline{w}} \|_2^2 - 2a\hat{\underline{w}}^T z^i \underline{x}^i + \| z^i \underline{x}^i \|_2^2 + 2\underline{w}(i)^T z^i \underline{x}^i$$

let $b^2 \triangleq \max_j \|\underline{x}_j\|_2^2$ = [length of ~~longest~~ data point (vector)]^2

$c \triangleq \min_j \{\hat{\underline{w}}^T z_j \underline{x}_j\} > m$

Then

$\|\underline{w}(i+1) - a\hat{\underline{w}}\|_2^2 = \|\underline{w}(i) - a\hat{\underline{w}}\|_2^2 - 2ac + b^2 + 2\underline{w}(i)^T z^i \underline{x}^i$

since $\underline{w}(i)^T z^i \underline{x}^i \le m$

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

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

since $a\hat{\underline{w}}^T z_n \underline{x}_n > m$ and $\hat{\underline{w}}^T z_n \underline{x}_n > m$

so $\frac{m}{\hat{\underline{w}}^T z_n \underline{x}^n} < 1$ and $a > \frac{m}{\hat{\underline{w}}^T z_n \underline{x}_n}$

so we just need to let $a \ge 1$

Now, choose $a = \frac{b^2 + 1}{c}$

$\varepsilon\, \underline{w}(i+1) \le \varepsilon\, \underline{w}(i) - b^2 + 2(m - c)$

since $c \triangleq \min_j \{\hat{\underline{w}}^T z_j \underline{x}_j\} > m$

so $m - c < 0$

so we have $\varepsilon\, \underline{w}(i+1) \le \varepsilon\, \underline{w}(i) - b^2$

so, each iteration reduces ε_w by at least b^2

Apply forcing argument;

$0 \le \varepsilon\, \underline{w}(i+1) \le \varepsilon\, \underline{w}(i) - b^2 \quad \forall i$

for some i_0, we would have : $\sum w(i_0) < b^3$

so that:

$$0 < \sum w(i_0+1) \leq \sum w(i_0) - b^2 < 0$$

$\rightarrow$ impossible

$\Rightarrow$ iterations must cease at $i = i_0 - 1$ (or sooner)

$\Rightarrow$ algorithm converges at a solution weight vector at (i_0-1) th iteration or soner

Problem2

2. first using augmented space

$$\begin{pmatrix} 1 \\ 0 \\ 1 \\ -1 \\ 2 \end{pmatrix} \quad \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad \begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad \begin{pmatrix} 1 \\ -1 \\ 1 \\ 6 \\ -1 \end{pmatrix}$$

$n(i) = 1$

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

$g_k(\underline{x}) = \underline{w}_k^T \underline{x}$

Decision rule $g_k(\underline{x}) > g_j(\underline{x}) \ \forall j \neq k \Rightarrow \underline{x} \in S_k$

$$\begin{pmatrix} 1 \\ 0 \\ 1 \\ -1 \\ 2 \end{pmatrix}$$

$g_1(\underline{x}^{(1)}) = -3$

$g_2(\underline{x}^{(1)}) = 3$

$g_3(\underline{x}^{(1)}) = 0$

$\underline{w}^{(1)}(1) = \begin{pmatrix} 0 \\ -1 \\ 0 \\ -2 \\ 1 \end{pmatrix}$

$\underline{w}^{(2)}(1) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 2 \\ -1 \end{pmatrix}$

$\underline{w}^{(3)}(1) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$

$$\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$g_1(\underline{x}^{(2)}) = -2$

$g_2(\underline{x}^{(2)}) = 2$

$g_3(\underline{x}^{(2)}) = 0$

$\underline{w}^{(1)}(2) = \begin{pmatrix} 0 \\ -1 \\ 0 \\ -2 \\ 1 \end{pmatrix}$

$\underline{w}^{(2)}(2) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 2 \\ -1 \end{pmatrix}$

$\underline{w}^{(3)}(2) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$

$$\begin{pmatrix}1\\2\\1\\1\\1\end{pmatrix} \quad g_1(x^{(3)}) = -3,\quad g_2(x^{(3)}) = 3,\quad g_3(x^{(3)}) = 0$$

$$\begin{pmatrix}1\\-1\\1\\0\\-1\end{pmatrix} \quad g_1(x^{(4)}) = 0,\quad g_2(x^{(4)}) = 0,\quad g_3(x^{(4)}) = 0$$

$$\underline{w}^{(1)}(3) = \begin{pmatrix}0\\-1\\0\\-2\\1\end{pmatrix} \qquad \underline{w}^{(1)}(4) = \begin{pmatrix}0\\-1\\0\\-2\\1\end{pmatrix}$$

$$\underline{w}^{(2)}(3) = \begin{pmatrix}0\\1\\0\\2\\-1\end{pmatrix} \qquad \underline{w}^{(2)}(4) = \begin{pmatrix}-1\\2\\-1\\2\\0\end{pmatrix}$$

$$\underline{w}^{(3)}(3) = \begin{pmatrix}0\\0\\0\\0\\0\end{pmatrix} \qquad \underline{w}^{(3)}(4) = \begin{pmatrix}1\\-1\\1\\0\\-1\end{pmatrix}$$

$$\begin{pmatrix}1\\0\\1\\-1\\2\end{pmatrix} \quad g_1(x^{(1)}) = 5,\quad g_2(x^{(1)}) = -4,\quad g_3(x^{(1)}) = 0$$

$$\begin{pmatrix}1\\1\\1\\1\\1\end{pmatrix} \quad g_1(x^{(2)}) = -2,\quad g_2(x^{(2)}) = 2,\quad g_3(x^{(3)}) = 0$$

$$\begin{pmatrix}1\\2\\1\\1\\1\end{pmatrix} \quad g_1(x^{(3)}) = -2,\quad g_2(x^{(3)}) = 3,\quad g_3(x^{(3)}) = -1$$

$$\begin{pmatrix}1\\-1\\1\\0\\-1\end{pmatrix} \quad g_1(x^{(4)}) = -4,\quad g_2(x^{(4)}) = 0,\quad g_3(x^{(4)}) = 4$$

$$\underline{w}^{(1)}(5) = \begin{pmatrix}0\\-1\\0\\-2\\1\end{pmatrix} \quad \underline{w}^{(1)}(6) = \begin{pmatrix}0\\-1\\0\\-2\\1\end{pmatrix} \quad \underline{w}^{(1)}(7) = \begin{pmatrix}0\\-1\\0\\-2\\1\end{pmatrix} \quad \underline{w}^{(1)}(8) = \begin{pmatrix}0\\-1\\0\\-2\\1\end{pmatrix}$$

$$\underline{w}^{(2)}(5) = \begin{pmatrix}-1\\2\\-1\\2\\0\end{pmatrix} \quad \underline{w}^{(2)}(6) = \begin{pmatrix}-1\\2\\-1\\2\\0\end{pmatrix} \quad \underline{w}^{(2)}(7) = \begin{pmatrix}-1\\2\\-1\\2\\0\end{pmatrix} \quad \underline{w}^{(2)}(8) = \begin{pmatrix}-1\\2\\-1\\2\\0\end{pmatrix}$$

$$\underline{w}^{(3)}(5) = \begin{pmatrix}1\\-1\\1\\0\\-1\end{pmatrix} \quad \underline{w}^{(3)}(6) = \begin{pmatrix}1\\-1\\1\\0\\-1\end{pmatrix} \quad \underline{w}^{(3)}(7) = \begin{pmatrix}1\\-1\\1\\0\\-1\end{pmatrix} \quad \underline{w}^{(3)}(8) = \begin{pmatrix}1\\-1\\1\\0\\-1\end{pmatrix}$$

Now, all data points are all correctly classified

So, we can know that

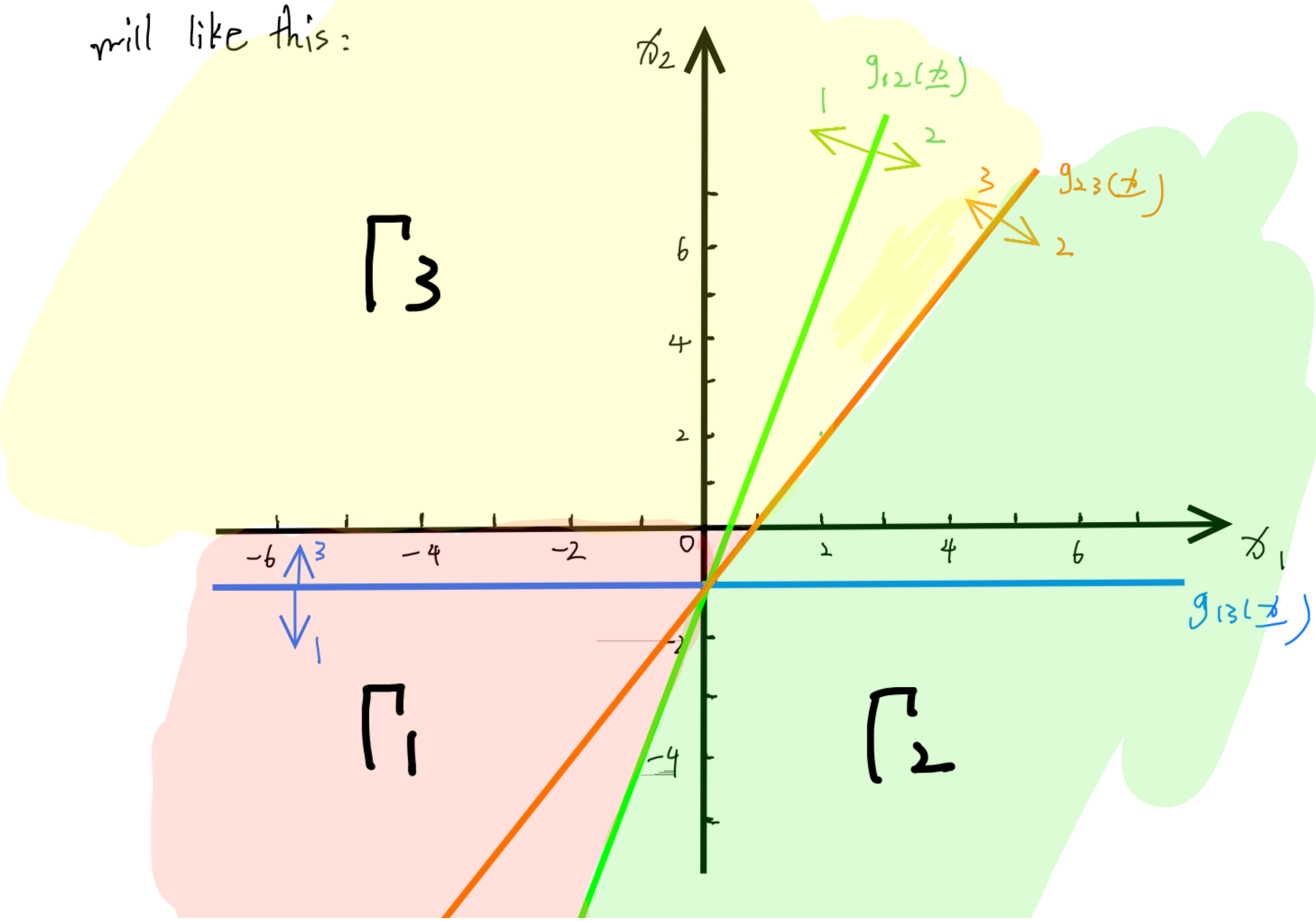
$$
\begin{cases}
g_1(\underline{x}) = -x_1 - 2x_3 + x_4 \\
g_2(\underline{x}) = -1 + 2x_1 - x_2 + 2x_3 \\
g_3(\underline{x}) = 1 - x_1 + x_2 - x_4.
\end{cases}
$$

(b)

so, using MVM, we can know that

$$
\begin{cases}
g_{12}(\underline{x}) \Rightarrow & x_2 = 3x_1 - 1 \\
g_{13}(\underline{x}) \Rightarrow & x_2 = -1 \\
g_{23}(\underline{x}) \Rightarrow & 2x_2 = 3x_1 - 2
\end{cases}
\qquad
\begin{cases}
x_2 - 3x_1 + 1 \\
x_2 + 1 \\
2x_2 - 3x_1 + 2
\end{cases}
$$

so, in 2D space, the decision boundaries and decision decision regions will like this:



Problem3

3.

(a) $J(\underline{w}) = \sum_{n=1}^{N} [\underline{w}^T z_n \underline{x}_n - b_n]^2 \quad b_n > 0$

so, $J_n(\underline{w}) = [w^T z_n x_n - b]^2 \quad b > 0$

let $\underline{x}_n = z_n x_n$

$\nabla_{\underline{w}} J_n(\underline{w}) = \nabla_{\underline{w}} [w^T \underline{x}^T \underline{x} w - 2 \underline{w}^T \underline{x}^T b_n + b_n^2]$

$\nabla_{\underline{w}} J_n(\underline{w}) = 2 \underline{x}_n^T \underline{x}_n w - 2 \underline{x}_n^T b_n$

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

(b)

begin initialize $\underline{w}(0) = 0$, $\underline{b}$, criterion θ, $\eta(1) = 1$, $k = 0$

random shuffle $\underline{x}$

For n in $[1, 2 \ldots N]$

let m be epoch

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

$\eta(k) = \frac{\eta(1)}{k}$

let $\eta'(k) = 2\eta(k)$

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

until halting

return $w(k+1)$