Homework 1

1a.

Misclassified, $\hat{y}(t) \neq y(t)$

$\downarrow$

$\mathrm{sign}(w(t)^T x(t))$

$y(t)$, +ve
$w(t)^T x(t)$, −ve
$y(t)$, −ve
$w(t)^T x(t)$, +ve

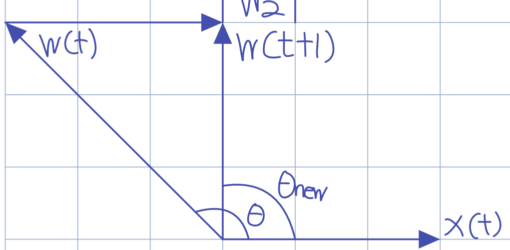
$y(t)\, w(t)^T x(t) < 0$

1b. $w(t+1) = w(t) + y(t)x(t)$

$$
\begin{aligned}
y(t)\, w(t+1)^T x(t) &= y(t)[w(t) + y(t)x(t)]^T x(t) \\
&= y(t)[w(t)^T + y(t)x(t)^T]x(t) \\
&= y(t)w(t)^T x(t) + y(t)y(t)x(t)^T x(t) \\
&> y(t)w(t)^T x(t)
\end{aligned}
$$

1c. $\hat{y}(t) = \mathrm{sign}(w(t)^T x(t))$

$$
w(t) = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}, \quad x(t) = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}
$$



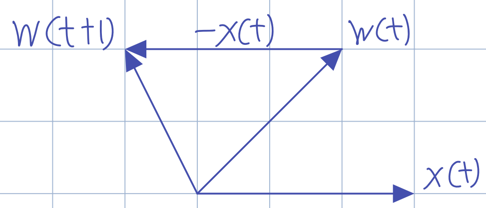
$$
\begin{aligned}
w_0 + w_1 x_1 + w_2 x_2 &= w(t) \cdot x(t) \\
&= |w(t)|\,|x(t)| \cos\theta
\end{aligned}
$$

$y(t) = +1$
$\hat{y}(t) = -1$

$w(t+1) = w(t) + y(t)x(t)$
Step in the right direction

$y(t) = -1$

$\hat{y}(t) = +1$

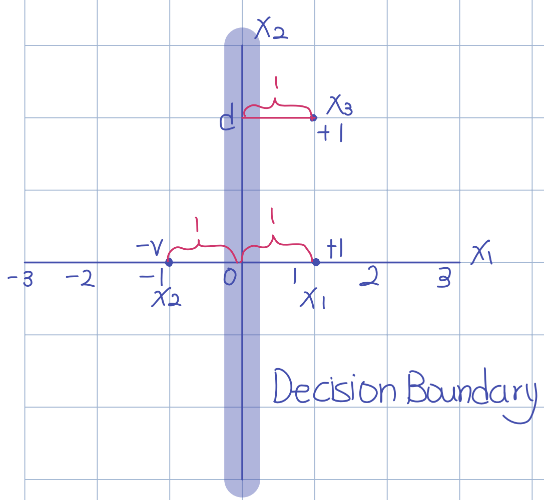


$w(t)x(t) = |w(t)|\,|x(t)|\cos\theta$

$w(t+1) = w(t) + y(t)x(t)$

$\qquad\quad w(t) - x(t)$

2a.



Yes, linearly separable

2b. $x_1 = (1, 0)$ $\quad x_2 = (-1, 0)$ $\quad x_3 = (1, d)$

$X_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad X_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad X_3 = \begin{bmatrix} 1 \\ 1 \\ d \end{bmatrix}$

Step 1, $w(0)^T x_1 = (0)(1) + (0)(1) + (0)(0) = 0$

$$w(1) = w(0) + y(t)x_1(t) = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

Step 2, $w(1)^T x_2 = (1)(1) + (1)(-1) + (0)(0) = 0$

$$w(2) = w(1) + y_2(t)x_2(t) = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$$

Step 3, $w(2)^T x_3 = (0)(1) + (2)(1) + (0)(d) = 2 > 0$ ✓

Step 4, $w(2)^T x_1 = (0)(1) + (2)(1) + (0)(0) = 2 > 0$ ✓

Step 5, $w(2)^T x_2 = (0)(1) + (2)(-1) + (0)(0) = -2 < 0$ ✓

$$\underline{w^*} = w(2) = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$$

$$w(2)^T x = \begin{bmatrix} 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} = 0 + 2x_1 + 0x_2 = 0$$

$$x_1 = 0$$

(8)

$$\text{margin} = 1 = \left| \frac{w^{*T} x}{\| w^* \|} \right|$$

2c)

Step 1, $w(0)^T x_3 = (0)(1)+(0)(1)+(0)(d)=0$ misclassified

$$w(1) = w(0) + y_3(t)x_3(t) = \begin{bmatrix}1\\1\\d\end{bmatrix}$$

Step 2, $w(1)^T x_2 = (1)(1)+(1)(-1)+(d)(0) = 0$ misclassified

$$w(2) = w(1) + y_2(t)x_2(t) = \begin{bmatrix}0\\2\\d\end{bmatrix}$$

Step 3, $w(2)^T x_1 = (0)(1)+(2)(1)+(d)(0) = 2>0$ ✓

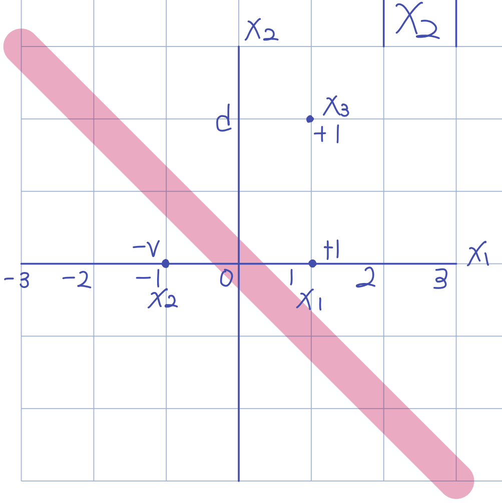
Step 4, $w(2)^T x_3 = (0)(1)+(2)(1)+(d)(d) = 2+d^2>0$ ✓

Step 5, $w(2)^T x_2 = (0)(1)+(2)(-1)+(d)(0) = -2<0$ ✓

$$\underline{w^*} = \begin{bmatrix}0\\2\\d\end{bmatrix}$$

$$w^{*T}x = \begin{bmatrix}0 & 2 & d\end{bmatrix}\begin{bmatrix}1\\x_1\\x_2\end{bmatrix} = 2x_1 + dx_2 = 0$$

$$x_1 = -\frac{d}{2}x_2$$



$$\rho_1 = \frac{w^{*T}x_1}{\|w^*\|} = \frac{2}{\sqrt{d^2+4}}$$

$$\rho_2 = \frac{w^{*T}x_2}{\|w^*\|} = \frac{2}{\sqrt{d^2+4}}$$

$$\rho_3 = \frac{w^{*T} x_3}{\|w^*\|} = \frac{2+d^2}{\sqrt{d^2+4}}$$

d $\uparrow$ slope $\downarrow$

2d,

$\rho_b = 1$ ✓

$$T \propto \frac{R^2}{\rho}$$

$$\rho_c = \frac{2}{\sqrt{4+d^2}} \approx 1$$ ✗

R, Radius of dataset

$\max(\|x\|^2)$

---

1. Linear Regressions

$y \in \mathbb{R}$

$$\mathcal{X} = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix} \in \mathbb{R}^{N\times d}, \quad \underline{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \in \mathbb{R}^N, \quad \underline{W} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix} \in \mathbb{R}^{d+1}$$

$$\mathcal{X}_{aug} = \begin{bmatrix} 1 & x_1^T \\ 1 & x_2^T \\ \vdots & \vdots \\ 1 & x_N^T \end{bmatrix} \in \mathbb{R}^{N\times(d+1)}$$

$$\hat{y} = w^T x = h(x)$$

Loss : Mean Squared Error

$$L_{sq} = (h(x) - y)^2$$

$$E_{in} = \frac{1}{2N} \sum_{n=1}^{N} (w^T x_n - y_n)^2$$

$$= \frac{1}{2N} \| \mathcal{X} \underline{w} - \underline{y} \|^2$$

$$\mathcal{X} \underline{w} = \begin{bmatrix} 1 & x_1^T \\ 1 & x_2^T \\ \vdots & \vdots \\ 1 & x_N^T \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix} = \begin{bmatrix} x_1^T & w_0 \\ x_2^T & w_1 \\ \vdots & \vdots \\ x_N^T & w_d \end{bmatrix}$$

Let $g = \underline{w^T \alpha} = \alpha_0 w_0 + \alpha_1 w_1 + \dots$

$$\frac{\partial g}{\partial w_0} = \alpha_0, \quad \frac{\partial g}{\partial w_1} = \alpha_1, \quad \frac{\partial g}{\partial w} = \underline{\alpha}$$

For $g = \underline{\alpha^T w}$

$$g = \underline{w^T} \, A \, \underline{w}$$

Matrix

$$\frac{\partial g}{\partial w} = (A + A^T) \underline{w}$$

If A Symmetric $\rightarrow A = A^T$

$$\frac{\partial g}{\partial w} = 2A\underline{w}$$

$$E_{in} = \frac{1}{2N}(\mathcal{X}\underline{w} - \underline{y})^T(\mathcal{X}\underline{w} - \underline{y})$$

$$= \frac{1}{2N}(\underline{w}^T\mathcal{X}^T - \underline{y}^T)(\mathcal{X}\underline{w} - \underline{y})$$

$$= \frac{1}{2N}\left[\underline{w}^T\mathcal{X}^T\mathcal{X}\underline{w} - \underline{w}^T\mathcal{X}^T y - \underline{y}^T\mathcal{X}\underline{w} + \underline{y^T y}\right]$$

$$\underline{y^T\mathcal{X}}\,\underline{w} = (\underline{w}^T\mathcal{X}^T y)^T \qquad = \frac{1}{2N}\left[\underline{w}^T\mathcal{X}^T\mathcal{X}\underline{w} - 2\underline{y}^T\mathcal{X}\underline{w} + y^T y\right]$$

$$\underline{w} \longrightarrow ((d+1)\times 1)$$

$$\underline{w}^T \longrightarrow (1\times(d+1))$$

$$\mathcal{X} \longrightarrow (N\times(d+1))$$

$$\mathcal{X}^T \longrightarrow ((d+1)\times N)$$

$$\underline{y} \longrightarrow (N\times 1)$$

$$\underline{w}^T\mathcal{X}^T\underline{y}$$

$$[1\times(d\times 1)][(d+1)\times N][N\times 1] = 1\times 1$$

$$\frac{\partial E_{in}}{\partial w} = 0 = \frac{1}{2N}\left[2\mathcal{X}^T\mathcal{X}\underline{w} - 2\underline{y}^T\mathcal{X}\right]$$

$$0 = 2\mathcal{X}^T\mathcal{X}\underline{w} - 2\underline{y}^T\mathcal{X}$$

$$\mathcal{X}^T\mathcal{X}\underline{w} = y^T\mathcal{X}$$

$$\underline{w} = (\mathcal{X}^T\mathcal{X}^{-1})\,\underline{y}^T\,\mathcal{X}$$

$x_1 = -1, x_2 = 0, x_3 = 1, x_4 = 1$

$\hat{x}_i = w_0 + w_1 x_{i-1}$

$x_{i-1} \longrightarrow x_i$

$x_1 \longrightarrow \hat{x}_2 \quad , \quad x_2 \longrightarrow \hat{x}_3 \quad , \quad x_3 \longrightarrow \hat{x}_4$

$$\mathcal{X} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \qquad y = \begin{bmatrix} x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

$$\mathcal{X}_{aug} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \qquad \hat{x}_i = w_0 + w_1 x_{i-1}$$

$$\underline{w}^* = (\mathcal{X}^T \mathcal{X})^{-1} \mathcal{X}^T \underline{y}$$

$$\mathcal{X}^T \mathcal{X} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$$

$$(\mathcal{X}^T \mathcal{X})^{-1} = \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

$$\mathcal{X}^T \underline{y} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\underline{w}^* = \begin{bmatrix} \frac{2}{3} \\ \frac{1}{2} \end{bmatrix}$$

$$f(w) = \log\left(\sum_{i=1}^{d} e^{w_i}\right)$$

$$\frac{\partial f(w)}{\partial w_1} = \frac{1}{\sum_{i=1}^{d} e^{w_i}} e^{w_1}$$

Chain Rule

$$\frac{\partial f}{\partial w_1} = \frac{1}{\sum_{i=1}^{d} e^{w_i}} \begin{bmatrix} e^{w_1} \\ \vdots \\ e^{w_d} \end{bmatrix}$$