HW2 - TRI VU   BME 590L

Problem 1.

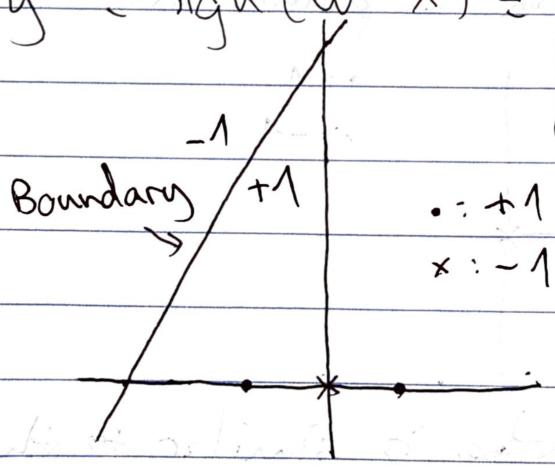
a) $X = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}$ (from Matlab) $\rightarrow X^+ = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ -1/2 & 0 & 1/2 \\ 0 & 0 & 0 \end{bmatrix}$

$y = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$

$w = X^+ y = \begin{bmatrix} 5/6 \\ 1/3 \\ -1/6 \end{bmatrix}$ (from Matlab)

b) $y^* = \text{sign}(w^T \hat{X})$ $\hat{X} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

$\rightarrow y^* = \text{sign}(w^T \hat{X}) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ $\Rightarrow$ the model classifies all points as +1



By looking at the data distribution, we can see that there is no possible ~~plan~~ linear planes solution to well-~~spe~~ separate the example points, i.e. all three points are linearly dependent

Proof: there is $[x\ y\ z] = [1\ 1\ 1]$ where $x\begin{bmatrix} -1 \\ 0 \end{bmatrix} + y\begin{bmatrix} 0 \\ 0 \end{bmatrix} + z\begin{bmatrix} 1 \\ 0 \end{bmatrix} = 0$

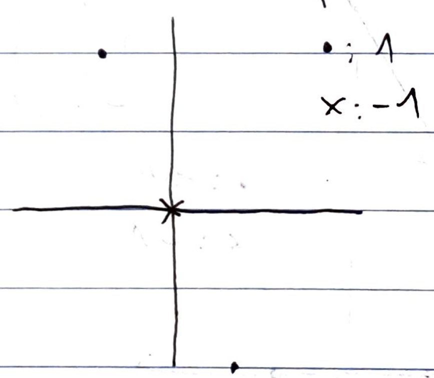
c, $W_1 = \begin{bmatrix} 3 & 0 & 0 \\ -1 & 3 & 0 \\ 0 & -1 & 3 \\ 0 & 0 & -1 \end{bmatrix}$ $\quad x_{pad} = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$

~~$W_1 x =$~~ We have x_{new}:

$$x_{new} = W_1 x_{pad} = \begin{bmatrix} 3 & 0 & 0 \\ -1 & 3 & 0 \\ 0 & -1 & 3 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\text{(from Matlab)} = \begin{bmatrix} 1 & -3 & 0 \\ 0 & 0 & 0 \\ -1 & 3 & 0 \end{bmatrix}$$

Since the third dimension is 0, we can ~~re~~ depict the mixed points on 2d plane as:



which ~~returns to the~~ is similar to the previous part that is non-separable by line.

(*) Mathematical proof

We have: $W_{new} = w_2^T W_1$

$\rightarrow W_{new} \in \mathbb{R}^{1\times 3}$

$\rightarrow y^* = \text{sign}(w_2^T W_1 x) = \text{sign}(W_{new} x)$ which goes back ~~#~~ to the steps of finding $\overset{w_{new}}{\cancel{w}} = X^{+} y$

→ no possible linear solution for $\underbrace{y^{*} = \mathrm{sign}(w_2^{T} W_1 x)}_{\text{classifier}}$

d, We have

$$W_1 \hat{X} = \begin{bmatrix} 3 & 0 & 0 \\ -1 & 3 & 0 \\ 0 & -1 & 3 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 3 & 3 \\ -4 & -1 & 2 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\rightarrow \mathrm{ReLU}(W_1 \hat{X}) = \begin{bmatrix} 3 & 3 & 3 \\ 0 & 0 & 2 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \cancel{x^{h}}\, x_c$$

$$\rightarrow \cancel{x_c^{+}}\; w_2 = x_c^{+} y = \begin{bmatrix} -1/3 \\ 1 \\ 2 \\ 0 \end{bmatrix}$$

$$y^{*} = \overset{\mathrm{sign}(w_2^{T} x_c)}{\cancel{w_2^{T} x_c}} = \cancel{\begin{bmatrix} -1/3 \\ 1 \\ 2 \\ 0 \end{bmatrix}} \begin{bmatrix} -1/3 & 1 & 2 & 0 \end{bmatrix} \begin{bmatrix} 3 & 3 & 3 \\ 0 & 0 & 2 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -1 & 1 \end{bmatrix}$$

→ correctly classified

Problem 2

a, $L_{in}(w) = \frac{1}{N}\sum_{n=1}^{N} \ln(1+e^{-y_n w^T x_n})$

$$\rightarrow \nabla_w L_{in}(w) = \frac{1}{N}\sum_{n=1}^{N} -y_n x_n \frac{e^{-y_n w^T x_n}}{1+e^{-y_n w^T x_n}}$$

with $\theta(-y_n w^T x_n) = \frac{e^{-y_n w^T x_n}}{1+e^{-y_n w^T x_n}}$

b, We have $\nabla_w L_{in}(w) \sim \theta(n)$ with $n = -y_n w^T x_n$

($n \in [-1, 1]$)

~~misclassified completely~~ ~~correctly classified~~

$$\theta(n) = \frac{e^n}{1+e^n} = 1 - \frac{1}{1+e^n}$$

* With $n \rightarrow 1$, e^n will increase from $0.368 \rightarrow 2.718$

$\rightarrow \frac{1}{1+e^n}$ will decrease and $1 - \frac{1}{1+e^n}$ will ~~decrease~~ increase

~~$\rightarrow \theta(n)$ will also decrease with increasingly correctly classified ($n \rightarrow 1$)~~

$\rightarrow \theta(n)$ will ~~de~~ increase when $n \rightarrow 1$, i.e. when the input is increasingly classified ($n = 1 \rightarrow y_n w^T x_n = -1$

(completely misclassified)

Problem 2c.

c) $$H_{Lin(w)} = \begin{bmatrix} \frac{\partial \nabla_w Lin(w)}{\partial x_n y_n} & \frac{\partial \nabla_w Lin(w)}{\partial x_n y_{n+1}} & \cdots \\ \frac{\partial \nabla_w Lin(w)}{\partial x_{n+1} y_n} & \frac{\partial \nabla_w Lin(w)}{\partial x_{n+1} y_{n+1}} & \\ \vdots & & \ddots \end{bmatrix}$$

From c), we eventually have:

$$H_{Lin(w)} = \begin{bmatrix} \frac{\partial \nabla_w Lin(w)}{\partial x_n y_n} & & 0 \\ & \frac{\partial \nabla_w Lin(w)}{\partial x_{n+1} y_{n+1}} & \\ 0 & & \ddots \end{bmatrix}$$

~~Opti~~ We also have Jacobian matrix as:

$$J_{\cancel{H}\,Lin(w)} = \begin{bmatrix} \frac{\partial Lin(w)}{\partial x_n y_n} & & 0 \\ & \frac{\partial Lin(w)}{\partial x_{n+1} y_{n+1}} & \\ 0 & & \ddots \end{bmatrix}$$

→ Optimal step size $\varepsilon = \frac{J^T J}{J^T H J}$

Problem 3.

a, Since $x(t)$ is misclassified by $w(t)$, we have:

$y(t) \neq \text{sign}(w^T(t)\,x(t))$

$\rightarrow$ $y(t)$ and $w^T(t)x(t)$ will have opposite sign

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

b,

$y(t)\,w^T(t+1)\,x(t)$

$= y(t)\,(w(t) + y(t)x(t))^T\,x(t)$

$= y(t)\,w^T(t)x(t) + \underbrace{y(t)x^T(t)\,y^T(t)\,x(t)}_{\|y \cdot x\|^2 > 0}$

$\rightarrow y(t)w^T(t+1)x(t) - y(t)\,w^T(t)x(t) > 0$

$\rightarrow y(t)\,w^T(t+1)\,x(t) > y(t)w^T(t)\,x(t)$ (1)

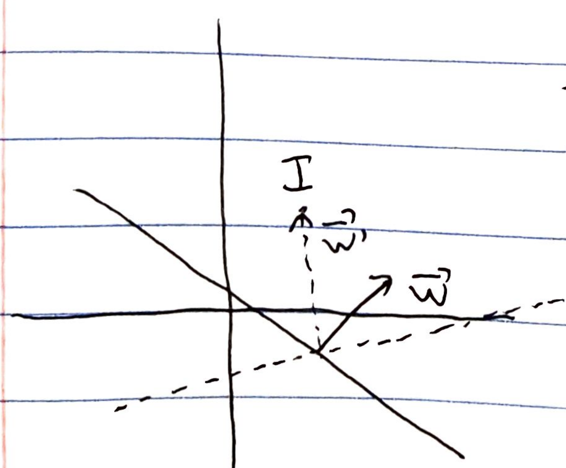
c, ~~Since~~ From (1), we have:

$w^T(t+1) \geq w^T(t)$

$\rightarrow w^T(t+1) - w^T(t) > 0$

~~$\rightarrow \nabla w > 0$~~

$\rightarrow$ $w(t)$ to $w(t+1)$ is the right direction



—— : boundary at $w(t)$

- - - : boundary at $w(t+1) = w'$

I : misclassified point

Problem 4.

First layer:

Input (zero-padded and reversed)

$$\begin{bmatrix} x(1)^1_1 \\ x(2)^1_1 \\ x(3)^1_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x(1)^1_1 \\ x(2)^1_1 \\ x(3)^1_1 \\ x(4)^1_1 \\ x(5)^1_1 \\ 0 \end{bmatrix} \quad \text{ReLU} \quad = b^1_1 + \begin{bmatrix} c(1)^1_1 & c(2)^1_1 & & & & 0 \\ & c(1)^1_1 & c(2)^1_1 & & & \\ & & c(1)^1_1 & c(2)^1_1 & & \\ & 0 & & c(1)^1_1 & c(2)^1_1 & \\ & & & & c(1)^1_1 & c(2)^1_1 \end{bmatrix} \begin{bmatrix} 0 \\ x_4 \\ x_3 \\ x_2 \\ x_1 \\ 0 \end{bmatrix}$$

bias

$$\begin{bmatrix} x(1)^1_2 \\ x(2)^1_2 \\ x(3)^1_2 \end{bmatrix} \qquad \begin{bmatrix} x(1)^1_2 \\ \cdots \\ x(5)^1_2 \\ 0 \end{bmatrix} \qquad b^1_2 + \begin{bmatrix} c(1)^1_2 & c(2)^1_2 & & & & 0 \\ & c(1)^1_2 & c(2)^1_2 & & & \\ & & c(1)^1_2 & c(2)^1_2 & & \\ & 0 & & c(1)^1_2 & c(2)^1_2 & \\ & & & & c(1)^1_2 & c(2)^1_2 \end{bmatrix}$$

zero-pad for sum-pooling

Sum-pooling layer

Convolutional layer

Second layer

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \overset{\text{ReLU}}{=} \begin{bmatrix} w(1)_1 & \cdots & w(8)_1 \\ w(1)_2 & \cdots & w(8)_2 \\ w(1)_3 & \cdots & w(8)_3 \end{bmatrix} \begin{bmatrix} x(1)^2_1 \\ x(2)^2_1 \\ x(1)^2_2 \\ x(2)^2_2 \\ x(1)^2_3 \\ x(2)^2_3 \\ x(1)^2_4 \\ x(2)^2_4 \end{bmatrix}$$

labels

$$\begin{bmatrix} x(1)^2_1 \\ x(2)^2_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x(1)^2_1 \\ x(2)^2_1 \\ x(3)^2_1 \\ x(4)^2_1 \end{bmatrix} \overset{\text{ReLU}}{=} b^2_1 + \begin{bmatrix} c(1)^2_1 & c(2)^2_1 & & 0 & \\ & c(1)^2_1 & c(2)^2_1 & & \\ & 0 & c(1)^2_1 & c(2)^2_1 & \\ & & & c(1)^2_1 & c(2)^2_1 \end{bmatrix} \begin{bmatrix} 0 \\ x(3)^1_1 \\ x(2)^1_1 \\ x(1)^1_1 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} x(1)^2_2 \\ x(2)^2_2 \end{bmatrix} \qquad \begin{bmatrix} x(1)^2_2 \\ x(2)^2_2 \\ x(3)^2_2 \\ x(4)^2_2 \end{bmatrix} \qquad b^2_2 + \begin{bmatrix} c(1)^2_2 & c(2)^2_2 & & 0 & \\ & c(1)^2_2 & c(2)^2_2 & & \\ & & c(1)^2_2 & c(2)^2_2 & \\ & 0 & & c(1)^2_2 & c(2)^2_2 \end{bmatrix} \begin{bmatrix} 0 \\ x(3)^1_2 \\ x(2)^1_2 \\ x(1)^1_2 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} x(1)^2_3 \\ x(2)^2_3 \end{bmatrix} \qquad \begin{bmatrix} x(1)^2_3 \\ x(2)^2_3 \\ x(3)^2_3 \\ x(4)^2_3 \end{bmatrix} \qquad \cancel{\begin{bmatrix} c(1)^2_3 & c(2)^2_3 & & & \\ & c(1)^2_3 & c(2)^2_3 & & \\ & & c(1)^2_3 & c(2)^2_3 & \\ & & & c(1)^2_3 & c(2)^2_3 \end{bmatrix}}$$

$$\begin{bmatrix} x(1)^2_4 \\ x(2)^2_4 \end{bmatrix} \qquad \begin{bmatrix} x(1)^2_4 \\ x(2)^2_4 \\ x(3)^2_4 \\ x(4)^2_4 \end{bmatrix} \qquad \cancel{\begin{bmatrix} c(1)^2_4 \end{bmatrix}}$$

FC layer — ~~Max~~ Sum-pooling layer — Convolutional layer