EXAM for

ARTIFICIAL NEURAL NETWORKS

CHALMERS, GÖTEBORGS UNIVERSITET

2018-10-29

1a) Define $Q^{(\mu,\nu)} = \sum_j x_j^{(\mu)} x_j^{(\nu)}$.

Find: $Q^{(1,1)} = \underline{\underline{32}}$

$Q^{(1,2)} = (32-13)-13$

$= \underline{\underline{6}}$ (13 bits differ)

$Q^{(1,3)} = 32-2-2$

$= \underline{\underline{28}}$ (2 bits differ)

$Q^{(1,4)} = -1 \cdot 32 = \underline{\underline{-32}}$ (inverse patterns)

$Q^{(1,5)} = \underline{\underline{0}}$ (orthogonal patterns)

Comparison of $\underline{x}^{(1)}$ and $\underline{x}^{(2)}$:

X	X	X	X
O	O	X	O
O	O	X	X
X	X	X	X
X	O	O	O
X	O	O	X
X	X	X	X
X	O	O	O

O: different bit

X: same bit

$Q^{(2,1)} = Q^{(1,2)} = \underline{\underline{6}}$ (symmetry)

$Q^{(2,2)} = \underline{\underline{32}}$

$Q^{(2,3)} = (32-15) - 15 = \underline{\underline{2}}$ (15 bits differ)

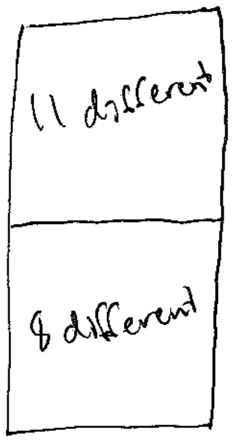
$Q^{(2,4)} = -Q^{(2,1)} = \underline{\underline{-6}}$ - (inverse pattern)

$Q^{(2,5)} = 32 - 19 - 19$

$= 32 - 38 = -6$

(19 bits different).

Comparision of $\underline{x}^{(2)}$ and $\underline{x}^{(5)}$:



(from previous fig.)

1b)

$$b_i^{(\nu)} = \sum_j w_{ij} x_j^{(\nu)}$$

$$= \sum_j \frac{1}{N}\left(x_i^{(1)} x_j^{(1)} + x_i^{(2)} x_j^{(2)}\right) x_j^{(\nu)}$$

$$= \frac{1}{N} x_i^{(1)} \sum_j x_j^{(1)} x_j^{(\nu)} + \frac{1}{N} x_i^{(2)} \sum_j x_j^{(2)} x_j^{(\nu)}$$

$$= \frac{1}{32} x_i^{(1)} Q^{(1,\nu)} + \frac{1}{32} x_i^{(2)} Q^{(2,\nu)}$$

From a):

$$b_i^{(1)} = \frac{1}{32} x_i^{(1)} Q^{(1,1)} + \frac{1}{32} x_i^{(2)} Q^{(2,1)}$$

$$= x_i^{(1)} + \frac{6}{32} x_i^{(2)}$$

$$b_i^{(2)} = \frac{1}{32} x_i^{(1)} Q^{(1,2)} + \frac{1}{32} x_i^{(2)} Q^{(2,2)} = \frac{6}{32} x_i^{(1)} + x_i^{(2)}$$

$$b_i^{(3)} = \frac{1}{32} x_i^{(1)} Q^{(1,3)} + \frac{1}{32} x_i^{(2)} Q^{(2,3)} = \frac{28}{32} x_i^{(1)} + \frac{2}{32} x_i^{(2)}$$

$$b_i^{(4)} = \frac{1}{32} x_i^{(1)} Q^{(1,4)} + \frac{1}{32} x_i^{(2)} Q^{(2,4)} = -x_i^{(1)} - \frac{6}{32} x_i^{(2)}$$

$$b_i^{(5)} = \frac{1}{32} x_i^{(1)} Q^{(1,5)} + \frac{1}{32} x_i^{(2)} Q^{(2,5)} = -\frac{6}{32} x_i^{(2)}$$

1c) Feeding $\underline{x}^{(\nu)}$, we find

$$S_i = \operatorname{sgn}\left[b_i^{(\nu)}\right]$$ after the update.

Pattern ν remains unchanged iff

$$x_i^{(\nu)} = \operatorname{sgn}\left[b_i^{(\nu)}\right] \text{ for all } i = 1, \ldots, 32.$$

$\underline{x}^{(1)}$: $\operatorname{sgn}\left[b_i^{(1)}\right] = x_i^{(1)}$: remains unchanged!

$\underline{x}^{(2)}$: $\operatorname{sgn}\left[b_i^{(2)}\right] = x_i^{(2)}$: remains unchanged!

$\underline{x}^{(3)}$: $\operatorname{sgn}\left[b_i^{(3)}\right] = x_i^{(1)}$: does not remain!

$\underline{x}^{(4)}$: $\operatorname{sgn}\left[b_i^{(4)}\right] = -x_i^{(1)} = x_i^{(4)}$: remains!

$\underline{x}^{(5)}$: $\operatorname{sgn}\left[b_i^{(5)}\right] = -x_i^{(2)}$: does not remain!

In summary patterns $\underline{x}^{(1)}$, $\underline{x}^{(2)}$ and $\underline{x}^{(4)}$ remains the same after update.

2a)

$$\sum_j w_{ij} x_j^{(mix)}$$

$$= \sum_j \frac{1}{N} \sum_\mu x_i^{(\mu)} x_j^{(\mu)} x_j^{(mix)}$$

$$= \sum_\mu x_i^{(\mu)} \frac{1}{N} \sum_j x_j^{(\mu)} x_j^{(mix)} = \sum_\mu x_i^{(\mu)} \langle s_\mu \rangle$$

2b)

$x_i^{(1)}$	$x_i^{(2)}$	$x_i^{(3)}$	$x_i^{(mix)}$	s_1	s_2	s_3
-1	-1	-1	-1	1	1	1
-1	-1	1	-1	1	1	-1
-1	1	-1	-1	1	-1	1
-1	1	1	1	-1	1	1
1	-1	-1	-1	-1	1	1
1	-1	1	1	1	-1	1
1	1	-1	1	1	1	-1
1	1	1	1	1	1	1

We find that $\langle s_1 \rangle$, $\langle s_2 \rangle$ and $\langle s_3 \rangle$ are $\frac{1}{N}$ times the sum of N independent random numbers X. These random numbers are such that

$$P[X = -1] = \frac{1}{4} \quad \text{and} \quad P[X = +1] = \frac{3}{4}.$$

Thus $E[X] = -\frac{1}{4} + \frac{3}{4} = \frac{1}{2}$

and $$\mathrm{Var}[X] = E\left[\left(X - \frac{1}{2}\right)^2\right]$$

$$= \frac{1}{4}\left(-\frac{3}{2}\right)^2 + \frac{3}{4}\left(\frac{1}{2}\right)^2$$

$$= \frac{1}{4}\,\frac{9}{4} + \frac{3}{4}\,\frac{1}{4} = \frac{12}{16} = \frac{3}{4}.$$

Thus $E[\langle s_\mu \rangle] = \frac{1}{2}$ for $\mu = 1, 2, 3$

and $\mathrm{Var}[\langle s_\mu \rangle] = \frac{1}{N^{\cancel{2}}} \cancel{N} \cdot \frac{3}{4} = \frac{3}{4N} \to 0$ as $N \to \infty$.

Thus $\langle s_\mu \rangle = \frac{1}{2}$ for $\mu = 1, 2, 3$.

For $\mu > 3$ we find that $\langle s_\mu \rangle$ is $\frac{1}{N}$ times the sum of N random numbers X, equal to ± 1 with equal probabilities.

Thus:

$E[X] = 0$, $Var[X] = 1$.

And, for $\mu > 3$:

$E[\langle s_\mu \rangle] = 0$ and $Var[\langle s_\mu \rangle]$

$= \frac{1}{N^2} N \cdot 1 = \frac{1}{N} \to 0$ as $N \to \infty$.

Thus $\langle s_\mu \rangle = 0$ for $\mu > 3$.

2c)

$$\sum_j w_{ij} x_j^{(mix)} = \sum_{\mu=1}^{3} x_i^{(\mu)} \underbrace{\langle s_\mu \rangle}_{=\frac{1}{2}} + \sum_{\mu>3} x_i^{(\mu)} \underbrace{\langle s_\mu \rangle}_{=0}$$

$$= \sum_\mu \frac{1}{2} x_i^{(\mu)} = \frac{1}{2}\left(x_i^{(1)} + x_i^{(2)} + x_i^{(3)} \right).$$

Cross-talk term can be neglected since $\langle s_\mu \rangle = 0$. $\langle s_\mu \rangle = 0$ follows from that it is $\frac{1}{N}$ times a sum of N terms, each with zero mean and unit variance (see 2b).

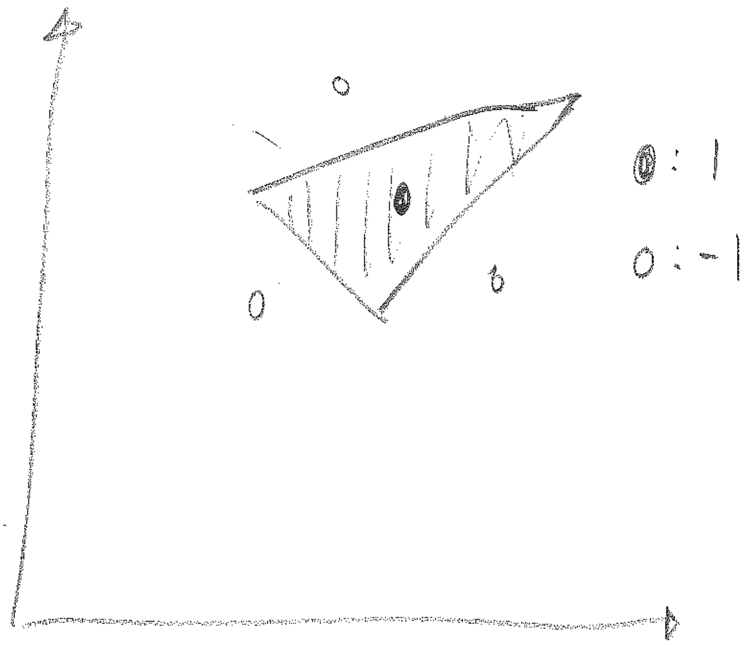
2d)

Apply update rule:

$$\mathrm{sgn}\left(\sum_j w_{ij} x_j^{(mix)}\right) = \{\text{from c})\}$$

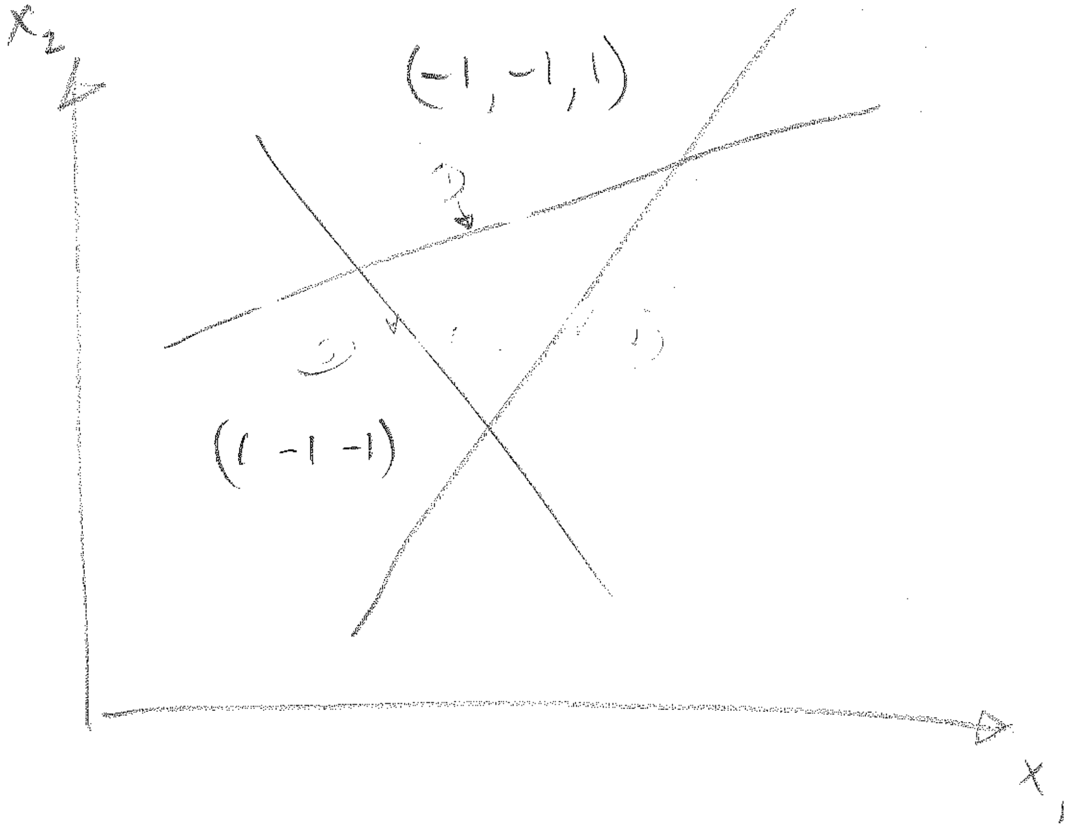
$$= \mathrm{sgn}\left(\frac{1}{2}\left(x_i^{(1)} + x_i^{(2)} + x_i^{(3)}\right)\right) = x_i^{(mix)}.$$

3. a)

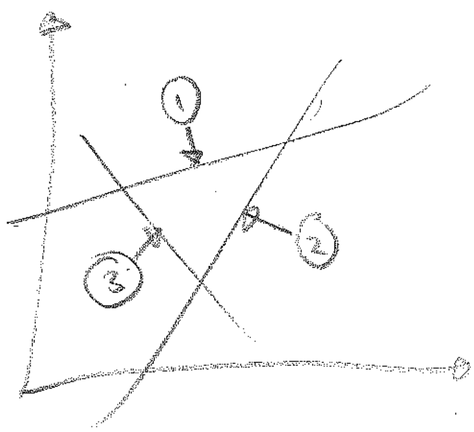


Not linearly separable.

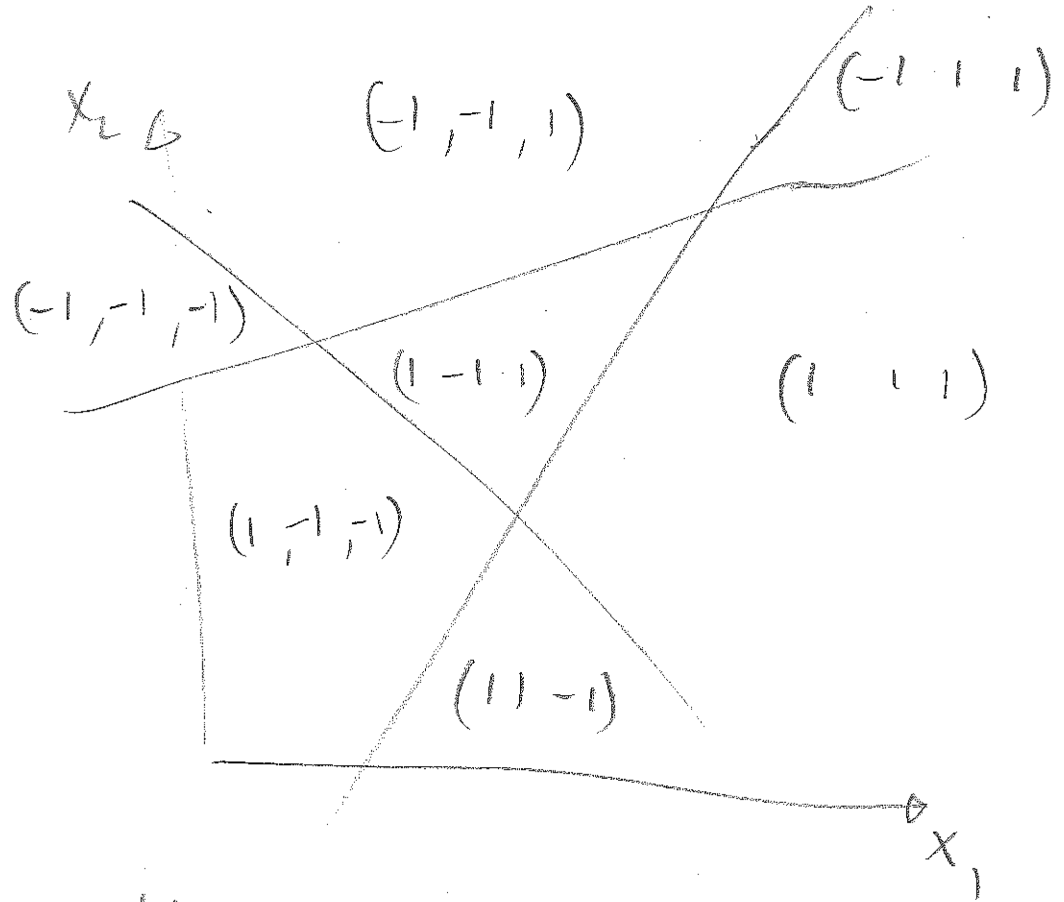
3 b) We are given



We deduce that the lines of the triangle must be ~~desis~~ decision boundaries of the (i)th hidden neuron according to:



Thus, the states of the hidden neurons are



Hidden neuron 1:

$\sum_k w_{1k} x_k - \theta_1 = 0$ for $\underline{x} = (-3, 2)$ and $\underline{x} = (2, 3)$

$$\Rightarrow \begin{cases} -3w_{11} + 2w_{12} - \theta_1 = 0 \\ 2w_{11} + 3w_{12} - \theta_1 = 0 \\ w_{11} = 1 \end{cases} \Rightarrow \begin{cases} w_{12} = -5 \\ \theta_1 = -13 \end{cases}$$

Hidden neuron 2:

$$\sum_k w_{2k} x_k - \Theta_2 = 0 \quad \text{for} \quad \underline{x} = (2,3) \text{ and } \underline{x} = (-1,-1)$$

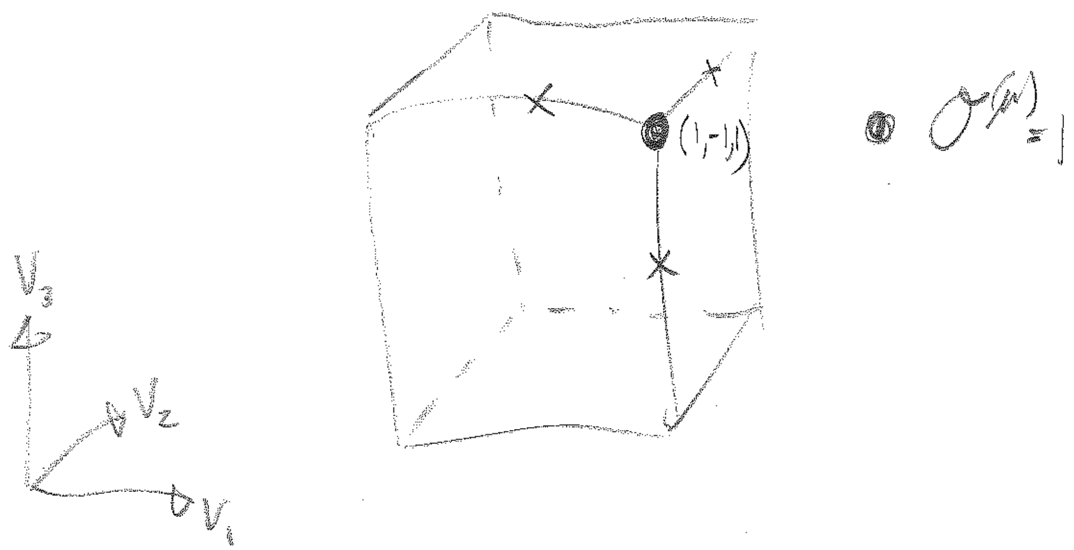
$$\Rightarrow \begin{cases} 2w_{21} + 3w_{22} - \Theta_2 = 0 \\ -w_{21} - w_{22} - \Theta_2 = 0 \\ w_{21} = 1 \end{cases} \Rightarrow \begin{cases} w_{22} = -\frac{3}{4} \\ \Theta_2 = -\frac{1}{4} \end{cases}$$

Hidden neuron 3

$$\sum_k w_{3k} x_k - \Theta_3 = 0 \quad \text{for} \quad \underline{x} = (-1,-1) \text{ and } \underline{x} = (-3,2)$$

$$\Rightarrow \begin{cases} -w_{31} - w_{32} - \Theta_3 = 0 \\ -3w_{31} + 2w_{32} - \Theta_3 = 0 \\ w_{31} = 1 \end{cases} \Rightarrow \begin{cases} w_{32} = \frac{2}{3} \\ \Theta_3 = -\frac{5}{3} \end{cases}$$

In order to find the W_j:s and Θ, we illustrate the space of the hidden neurons:



We want a DB passing through the crosses:

$$\sum_{j=1}^{3} W_j V_j - \Theta = 0 \quad \text{for}$$

$\underline{V} = (1, -1, 0)$,

$\underline{V} = (0, -1, 1)$ and

$\underline{V} = (1, 0, 1)$.

This gives:

$$\begin{cases} \overline{W}_1 - \overline{W}_2 - \Theta = 0 \\ -\overline{W}_2 + \overline{W}_3 - \Theta = 0 \\ \overline{W}_1 + \overline{W}_3 - \Theta = 0 \end{cases}$$

We need $\overline{W}_1 > 0$ (from figure),

Set $\underline{\overline{W}_1 = 1}$ and find:

$\underline{\overline{W}_2 = -1}$, $\underline{\overline{W}_3 = 1}$ and $\underline{\Theta = 2}$.

4)

The update formulae are

$$\begin{cases} w_{mn} \leftarrow w_{mn} + \delta w_{mn} \\ \Theta_m \leftarrow \Theta_m + \delta\Theta_m \\ W_{1m} \leftarrow W_{1m} + \delta W_{1m} \\ \Theta_1 \leftarrow \Theta_1 + \delta\Theta_1 \end{cases} \quad (1)$$

with

$$\begin{cases} \delta w_{mn} = -\eta \dfrac{\partial H^{(n)}}{\partial w_{mn}} \\ \delta\Theta_m = -\eta \dfrac{\partial H^{(n)}}{\partial \Theta_m} \\ \delta W_{1m} = -\eta \dfrac{\partial H^{(n)}}{\partial W_{1m}} \\ \delta\Theta_1 = -\eta \dfrac{\partial H^{(n)}}{\partial \Theta_1} \end{cases} \quad (2)$$

In ②, $H^{(\mu)} = \frac{1}{2}\left(t_1^{(\mu)} - O_1^{(\mu)}\right)^2$.

Define

$$\begin{cases} b_m^{(\mu)} = \sum_{n=1}^{3} w_{mn} x_n^{(\mu)} - \theta_m \\ B_1^{(\mu)} = \sum_{m=1}^{5} W_{1m} V_m^{(\mu)} - \Theta_1 . \end{cases}$$

Find:

$$\frac{\partial H^{(\mu)}}{\partial W_{1m}} = -\left(t_1^{(\mu)} - O_1^{(\mu)}\right) \frac{\partial}{\partial W_{1m}} g\left(B_1^{(\mu)}\right)$$

$$= \underbrace{\left(O_1^{(\mu)} - t_1^{(\mu)}\right) g'\left(B_1^{(\mu)}\right)}_{\equiv \Delta_1^{(\mu)}} \frac{\partial}{\partial W_{1m}} \left(\sum_{p=1}^{5} W_{1p} V_p^{(\mu)} - \Theta_1 \right)$$

$$= \Delta_1^{(\mu)} V_m^{(\mu)} \qquad (*)$$

$$\sum_p \delta_{pm} V_p$$

$$\frac{\partial H^{(\mu)}}{\partial \Theta_1} = -\Delta_1^{(\mu)} \qquad (*)$$

$$\frac{\partial H^{(\mu)}}{\partial w_{mn}} = \Delta_1^{(\mu)} \sum_{p=1}^{5} W_{1p} \frac{\partial V_p^{(\mu)}}{\partial w_{mn}}$$

$$= \Delta_1^{(\mu)} \sum_{p} W_{1p}\, g'(b_p^{(\mu)}) \frac{\partial}{\partial w_{mn}} \left(\sum_{q=1}^{3} w_{pq} x_q^{(\mu)} - \Theta_p \right)$$

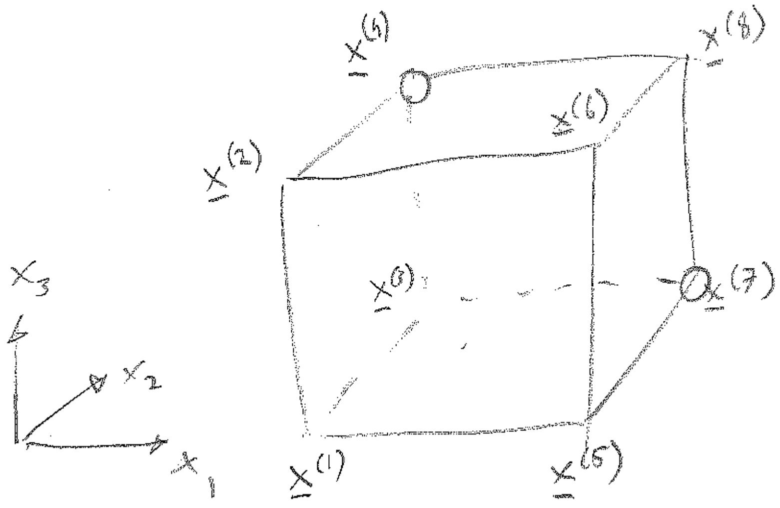
$$= \sum_p \underbrace{\Delta_1^{(\mu)} W_{1p}\, g'(b_p^{(\mu)})}_{\equiv\, \delta_p^{(\mu)}} \sum_q \delta_{pm} \delta_{nq} x_q^{(\mu)}$$

$$= \delta_m^{(\mu)} x_n^{(\mu)} \qquad (*)$$

$$\frac{\partial H^{(\mu)}}{\partial \Theta_m} = -\delta_m^{(\mu)}. \qquad (*)$$

Update rules are given by (1) with (2) and the (*):s inserted.

5a)

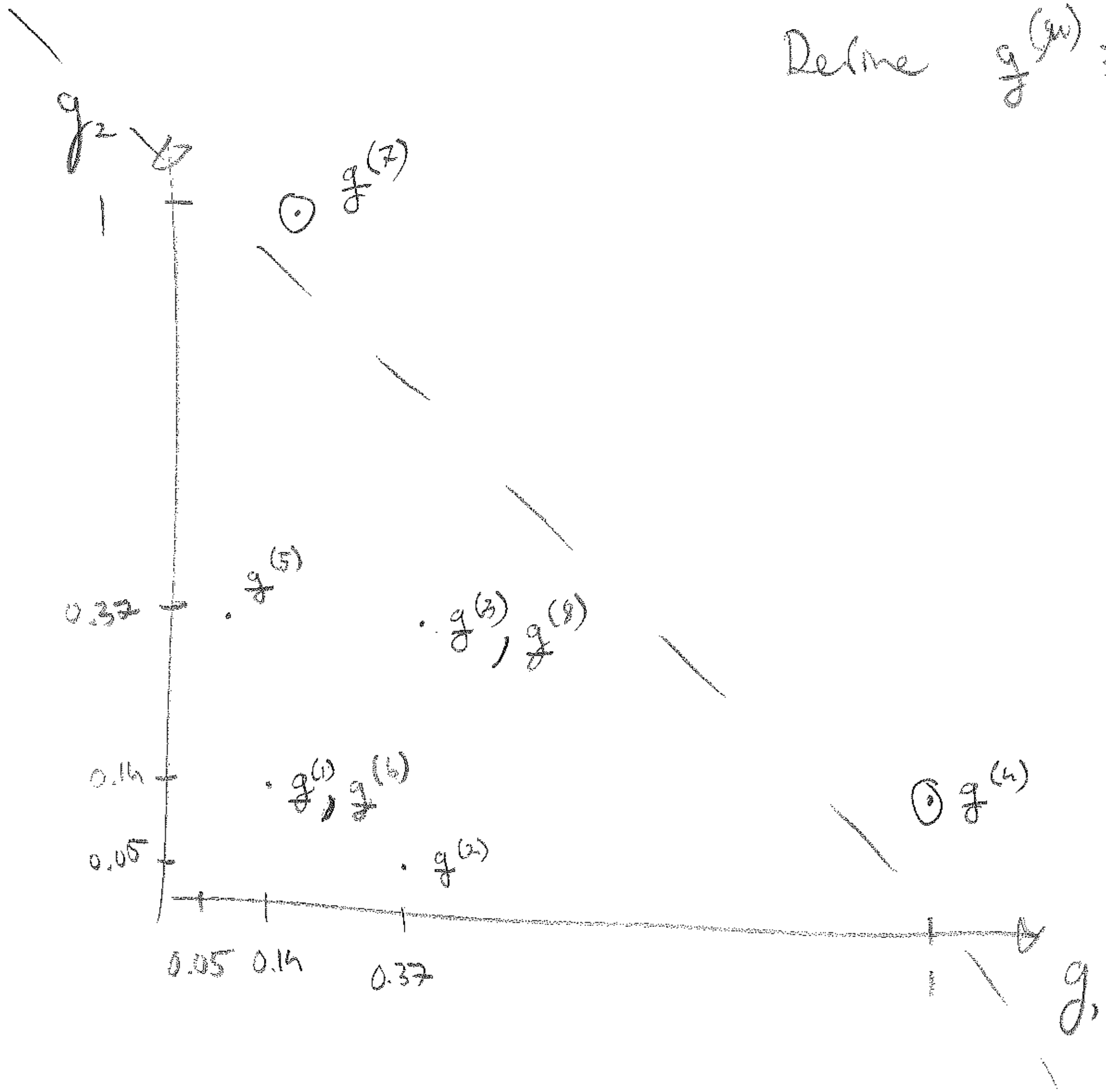


Can not be solved by the simple perceptron specified. Not linearly separable.

5b.1 and 5b.2

μ	$\underline{x}^{(\mu)}$	$\underline{x}^{(\mu)}-\underline{w}_1$	$\underline{x}^{(\mu)}-\underline{w}_2$	$\lvert\underline{x}^{(\mu)}-\underline{w}_1\rvert^2$	$\lvert\underline{x}^{(\mu)}-\underline{w}_2\rvert^2$	$g_1^{(\mu)}$	$g_2^{(\mu)}$
1	$(-1\ -1\ -1)$	$(0\ -2\ -2)$	$(-2\ -2\ 0)$	8	8	0.14	0.14
2	$(-1\ -1\ 1)$	$(0\ -2\ 0)$	$(-2\ -2\ 2)$	4	12	0.37	0.05
3	$(-1\ 1\ -1)$	$(0\ 0\ -2)$	$(-2\ 0\ 0)$	4	4	0.37	0.37
4	$(-1\ 1\ 1)$	$(0\ 0\ 0)$	$(-2\ 0\ 2)$	0	8	1	0.14
5	$(1\ -1\ -1)$	$(2\ -2\ -2)$	$(0\ -2\ 0)$	12	4	0.05	0.37
6	$(1\ -1\ 1)$	$(2\ -2\ 0)$	$(0\ -2\ 2)$	8	8	0.14	0.14
7	$(1\ 1\ -1)$	$(2\ 0\ -2)$	$(0\ 0\ 0)$	8	0	0.14	1
8	$(1\ 1\ 1)$	$(2\ 0\ 0)$	$(0\ 0\ 2)$	4	4	0.37	0.37

Define $\underline{g}^{(n)} \equiv \underline{g}(\underline{x}^{(n)})$



5.b.3-)

We have $W_1 g_1 + W_2 g_2 - \Theta = 0$ at $(g_1, g_2) = (1, 0)$ and at $(g_1, g_2) = (0, 1)$

This gives

$$\begin{cases} W_1 - \Theta = 0 \\ W_2 - \Theta = 0 \end{cases} \Rightarrow W_1 = W_2 = \Theta.$$

Output decreases when crossing the DB from below/left

$\Rightarrow W_2 < 0$ and $W_1 < 0$. Choose $W_1 = W_2 = \Theta = -1$.

6a)

Define $b_j = b_j^{(L,\mu)}$ and $S = \sum_m e^{b_m}$ and $\sigma_i = \sigma_i^{(\mu)}$

Find $\frac{\partial S}{\partial b_j} = \frac{\partial}{\partial b_j} \sum_m e^{b_m}$

$$= \sum_m \frac{\partial}{\partial b_j} e^{b_m} = \sum_m \delta_{jm} e^{b_j}$$

$$= e^{b_j}$$

Now find:

$$\frac{\partial \sigma_i}{\partial b_j} = \frac{\partial}{\partial b_j} e^{b_i} S^{-1} =$$

$$e^{b_i} \frac{\partial}{\partial b_j} S^{-1} + S^{-1} \frac{\partial}{\partial b_j} e^{b_i}$$

$$= -e^{b_i} S^{-2} \frac{\partial S}{\partial b_j} + S^{-1} e^{b_i} \delta_{ij}$$

$$= S^{-1} e^{b_i} \left(\delta_{ij} - S^{-1} e^{b_j} \right) = \sigma_i (\delta_{ij} - \sigma_j^{(\mu)})$$

$$\Rightarrow \frac{\partial \sigma^{(\mu)}}{\partial b_j^{(L,\mu)}} = \sigma_i^{(\mu)} (\delta_{ij} - \sigma_j^{(\mu)}).$$

6b)

$$\frac{\partial H}{\partial w_{pq}^{(L)}} =$$

$$= \frac{\partial}{\partial w_{pq}^{(L)}} - \sum_{i\mu} t_i^{(\mu)} \log O_i^{(\mu)}$$

$$= - \sum_{i\mu} t_i^{(\mu)} \frac{\partial}{\partial w_{pq}^{(L)}} \log O_i^{(\mu)}$$

$$= - \sum_{i\mu} t_i^{(\mu)} \frac{1}{O_i^{(\mu)}} \frac{\partial O_i^{(\mu)}}{\partial w_{pq}^{(L)}}$$

$$= - \sum_{i\mu} t_i^{(\mu)} \frac{1}{O_i^{(\mu)}} \sum_{j} \frac{\partial O_i^{(\mu)}}{\partial b_j^{(L,\mu)}} \frac{\partial b_j^{(L,\mu)}}{\partial w_{pq}^{(L)}}$$

$$= \{\text{from a)}\}$$

$$= -\sum_{i\mu} t_i^{(\mu)} \frac{1}{o_i^{(\mu)}} \sum_j o_i^{(\mu)} \left(\delta_{ij} - o_j^{(\mu)}\right) \frac{\partial b_j^{(L,\mu)}}{\partial w_{pq}^{(L)}}$$

$$= \sum_{i\mu} t_i^{(\mu)} \frac{1}{\cancel{o_i^{(\mu)}}} \sum_j \cancel{o_i^{(\mu)}} o_j^{(\mu)} \frac{\partial b_j^{(L,\mu)}}{\partial w_{pq}^{(L)}}$$

$$- \sum_{i\mu} t_i^{(\mu)} \frac{1}{\cancel{o_i^{(\mu)}}} \sum_j \cancel{o_i^{(\mu)}} \delta_{ij} \frac{\partial b_j^{(L,\mu)}}{\partial w_{pq}^{(L)}}$$

$$= \sum_{j\mu} o_j^{(\mu)} \frac{\partial b_j^{(L,\mu)}}{\partial w_{pq}^{(L)}} \cancel{\sum_i t_i^{(\mu)}} \to 1$$

$$- \sum_{i\mu} t_i^{(\mu)} \frac{\partial b_i^{(L,\mu)}}{\partial w_{pq}^{(L)}} = \sum_{i\mu} \left(o_i^{(\mu)} - t_i^{(\mu)}\right) \frac{\partial b_i^{(L,\mu)}}{\partial w_{pq}^{(L)}}$$

$$= \sum_{i,\mu} \left(O_i^{(\mu)} - t_i^{(\mu)} \right) \frac{\partial}{\partial w_{pq}^{(L)}} \sum_k w_{jk}^{(L)} V_k^{(L-1,\mu)}$$

$$= \sum_{i,\mu} \left(O_i^{(\mu)} - t_i^{(\mu)} \right) \sum_k \delta_{ip} \delta_{kq} V_k^{(L-1,\mu)}$$

$$= \sum_{\mu} \left(O_p^{(\mu)} - t_p^{(\mu)} \right) V_q^{(L-1,\mu)}$$

Thus:

$$\delta w_{pq}^{(L)} = -\eta \frac{\partial H}{\partial w_{pq}^{(L)}} = -\eta \sum_{\mu} \left(O_p^{(\mu)} - t_p^{(\mu)} \right) V_q^{(L-1,\mu)}$$