

Artificial Neural Networks Lecture Notes

Stephen Lucci, PhD

Part 2

About this file:

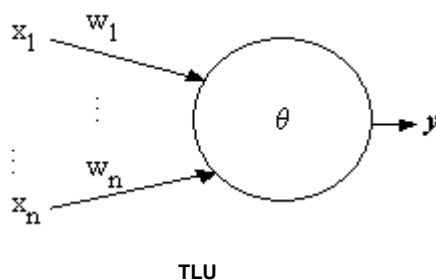
- If you have trouble reading the contents of this file, or in case of transcription errors, email gi0062@bcmail.brooklyn.cuny.edu
- Acknowledgments:
Background image is from <http://www.anatomy.usyd.edu.au/online/neuroanatomy/tutorial1/tutorial1.html> (edited) at the [University of Sydney Neuroanatomy web page](http://www.anatomy.usyd.edu.au/online/neuroanatomy/tutorial1/tutorial1.html). Mathematics symbols images are from [metamath.org](http://www.metamath.org)'s [GIF images for Math Symbols](http://www.metamath.org) web page. Other image credits are given where noted, the remainder are native to this file.

Contents

- Threshold Logic
 1. Threshold Logic Units
 2. Resilience To Noise and Hardware Failure
 3. Non-Linear Systems
 4. Geometric Interpretation of TLU action
 5. Vectors
 6. TLU's and Linear Separability Revisited

Threshold Logic

- Threshold Unit or Threshold Logic Unit (TLU)



▪ Activation $a = w_1 \cdot x_1 + w_2 \cdot x_2 + w_n \cdot x_n$

Example: $\vec{w} = (0.5, 1.0, 1.2)$,
 $\vec{x} = (1, 1, 0)$,

then $\vec{w} \cdot \vec{x} = (0.5 \cdot 1) + (1.0 \cdot 1) + (1.2 \cdot 0)$
 $= 0.5 + 1.0$
 $= 1.5$

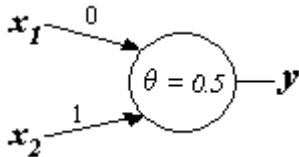
- To emulate the generation of action potentials, we need a threshold value θ .

$$y = \begin{cases} 1 & \text{if } a \geq \theta \\ 0 & \text{if } a < \theta. \end{cases}$$

- In the previous example, if $\theta = 1$, then y would equal 1 (because a is ≥ 1).

○ **Resilience To Noise and Hardware Failure**

- Consider the TLU shown below, and whose behavior is shown in the table to the right. The TLU fires, or outputs one only when x_2 is 1:



x_1	x_2	Activation	Output
0	0	0	0
0	1	1	1
1	0	0	0
1	1	1	1

- Now, suppose the hardware which implements the TLU is faulty, thus giving us weights encoded as (0.2, 0.8). The revised table shown above would then be as follows:

x_1	x_2	Activation	Output
0	0	0	0
0	1	0.8	1
1	0	0.2	0
1	1	1	1

- In the revised table above, we see that the activation has changed, but the output is the same. Thus:
- ***Changes in the activation, as long as they don't cross the threshold, produce no change in the output.***

○ **Non-Linear Systems**

- In a linear system, the output is proportionally related to the input:
i.e., small/large changes in the input always produce corresponding small/large changes in the output.
- Non-linear systems do not obey a proportionality restraint
so that the magnitude of change in output does not necessarily reflect that of the input.

Example:

In the above TLU, consider a case where an activation change from 0.0 to 0.2 produces no change in the output, whereas an activation change from 0.49 to 0.51 produces change in the output from 0 to 1. This is because 0.51 is larger than the threshold 0.5.

- Similarly, if input signals become degraded in some way, due to noise or a partial loss, once again, correct outputs occur.

Example:

instead of input value 1, we have 0.8, and instead of input 0, we have 0.2. Our revised table appears below:

x_1	x_2	Activation	Output
0.2	0.2	0.2	0
0.2	0.8	0.8	1
0.8	0.2	0.2	0
0.8	0.8	0.8	1

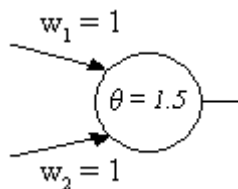
non-linear behavior of TLU's

- Thus, TLU's are robust in the presence of noisy or corrupted input signals.
- **Graceful degradation:** In a large network, as the degree of hardware and/or signal degradation increases, the number of TLU's giving incorrect results will gradually increase as well.

Contrast this with what happens in the case of *conventional* computers.

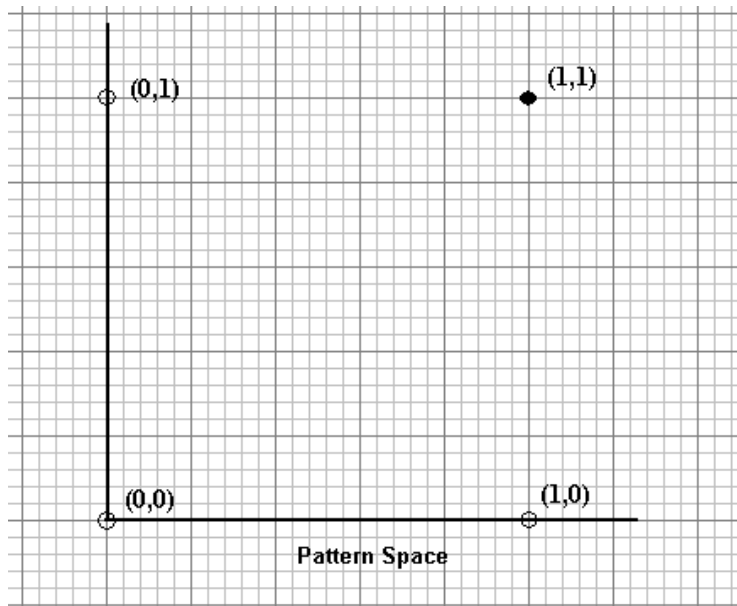
○ **Geometric Interpretation of TLU action**

- A TLU separates its input patterns into two categories according to its binary response (0 or 1) to each pattern.
- Consider the TLU shown here which classifies its input patterns into two groups: those that give output 0, and those that give output 1.



x_1	x_2	Activation	Output
0	0	0	0
0	1	1	0
1	0	1	0
1	1	2	1

- The **pattern space** of this TLU is as follows:



- **The Linear Separation of Classes**

Inputs: x_1, x_2

Activation: a

Threshold: θ .

- The critical condition for classification occurs when the activation equals the threshold
- we have:

$$w_1 \cdot x_1 + w_2 \cdot x_2 = \theta$$

Subtracting $w_1 \cdot x_1$ from both sides

$$w_2 \cdot x_2 = -w_1 \cdot x_1 + \theta$$

Dividing both sides by w_2 gives

$$x_2 = -(w_1 / w_2) x_1 + (\theta / w_2) .$$

This is of the general form:

$$y = m \cdot x + b$$

Recall that this is the linear (line) equation where m is the slope, and b is the y -intercept.

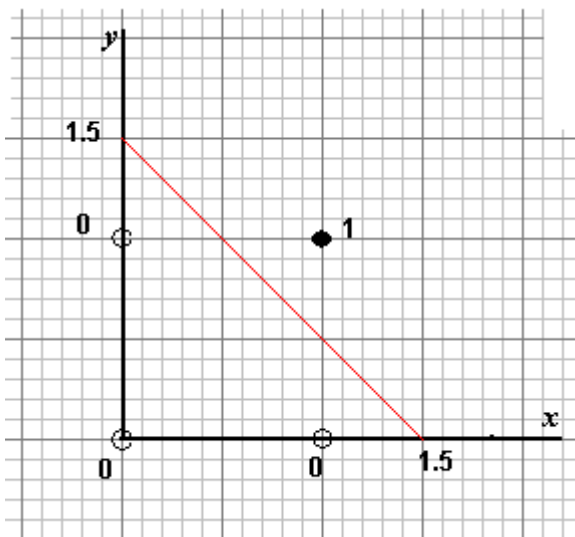
- Our previous example revisited:

$$w_1 = 1, w_2 = 1, \theta = 1.5.$$

$$\therefore m = -(w_1 / w_2) = -1 ,$$

$$b = (\theta / w_2) = 1.5 / 1 = 1.5 .$$

$$y = -x + 1.5$$



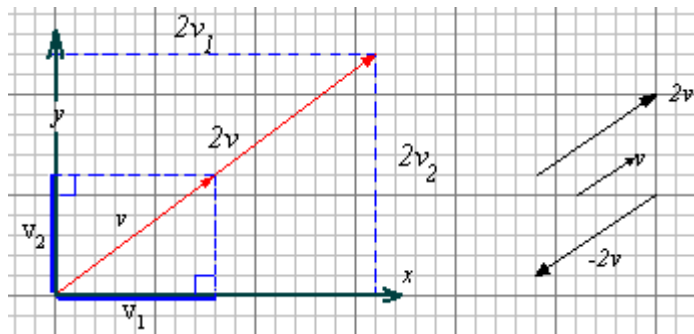
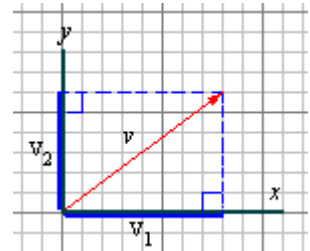
x	y
1.5	0
0	1.5

The two **classes** of the TLU output are separated by the red line.
 This will always be the case - in R^n we will have separating hyperplanes.

- **TLU's are linear separators and their patterns are linearly separable.**

○ **Vectors**

- A vector has a *magnitude* and a *direction*.
- We denote a vector v with \vec{v} , (in Europe, \underline{v} .)
- We denote a vector's magnitude with $\|\vec{v}\|$; sometimes with r .
- A vector is defined by a pair of numbers $(\|\vec{v}\|, \theta)$, where θ is the angle the vector makes with some reference direction (e.g., the x-axis).
- Alternate Representation - Cartesian co-ordinate system.
 For example, $\vec{v} = (v_1, v_2)$, where v_1, v_2 are the *components* of the vector.
- With the above representation, the vector can now be considered as an ordered pair, or more generally, an ordered list of numbers - note that the **order** is important. (e.g., in general, $(v_1, v_2) \neq (v_2, v_1)$.)
- Scalar Multiplication of a vector:



- Generalizing to n dimensions, we have $k \vec{v}$

$$= (kv_1, kv_2, \dots, kv_n).$$

- Vector Addition:

- Two vectors may be added in 2D by appending one to the end of the other. Note that the vector may be drawn anywhere in the space as long as its magnitude and direction are preserved.
- In terms of its components, if $\vec{w} = \vec{u} + \vec{v}$, then

$$w_1 = u_1 + v_1$$

$$w_2 = u_2 + v_2.$$
- In n -dimensions, we have:

$$\vec{w} = \vec{u} + \vec{v}$$

$$\vec{w} = (u_1 + v_1, u_2 + v_2, \dots, u_n + v_n).$$
- Note: Addition is commutative. (i.e., $\vec{u} + \vec{v} = \vec{v} + \vec{u}$.)

▪ Vector Substraction

▪ The Length of a Vector

- in 2D, $\|\vec{v}\| = \text{sqrt}(v_1^2, u_2^2)$.
- in n-dimensions, $\|\vec{v}\| = \text{sqrt}(\sum v_i^2)$, with i running from 1 to n.

▪ Comparing Vectors

▪ The Inner Product - Geometric Form:

- Useful to have some measure of how well aligned two vectors are. i.e., to know whether they point in the same or opposite direction.

▪ Inner Product:

$$\vec{v} \cdot \vec{w} = \|\vec{v}\| \|\vec{w}\| \cos\theta$$

where θ is the angle between the two.

This product is also known as the **scalar product**.

Note that if θ is $> 360^\circ$ we have equivalence to $360 - \theta$.

▪ The Inner Product - Algebraic Form:

- Given $\vec{v} = (1,1)$, $\vec{w} = (2,0)$, (note the angle between them is 45° .)

- observe that $\|\vec{v}\| = \text{sqrt}(1^2+1^2) = \text{sqrt}(2)$,
and that $\|\vec{w}\| = \text{sqrt}(2^2+0^2) = 2$.

$$\therefore \vec{v} \cdot \vec{w} = \text{sqrt}(2) * 2 * (1/\text{sqrt}(2)) = 2.$$

- Alternatively, we have

$$\begin{aligned} \vec{v} \cdot \vec{w} &= v_1 w_1 + v_2 w_2 \\ &= 1*2 + 1*0 = 2. \end{aligned}$$

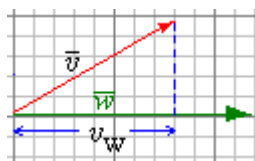
- This is also called the **dot product**.

- In n-dimensions, we have

$$\vec{v} \cdot \vec{w} = \sum v_i w_i, \text{ with } i \text{ running from } 1 \text{ to } n.$$

- If the sum is positive, then the vectors are lined up to some extent.
If the sum is zero, then the vectors are at right angles.
If the sum is negative, then the vectors are pointing away from each other.

Vector Projection



How much of \vec{v} lies in the direction of \vec{w} ?

Drop a perpendicular projection v_w of \vec{v} onto \vec{w} , where

$$v_w = \|\vec{v}\| \cos\theta.$$

- Alternatively, we may write:

$$\begin{aligned} v_w &= (\|\vec{v}\| \|\vec{w}\| \cos\theta) / \|\vec{w}\| \\ &= (\vec{v} \cdot \vec{w}) / \|\vec{w}\|. \end{aligned}$$

○ **TLU's and Linear Separability Revisited**

- The activation a of an n-input TLU is given by

$$a = \vec{w} \cdot \vec{x}.$$

- What happens when the activation equals the threshold?

Let $n = 2$. We have

$$w_1x_1 + w_2x_2 = \theta$$

i.e., $\bar{w} \cdot \bar{x} = \theta$ (*)

- For an arbitrary \bar{x} , the projection of \bar{x} into \bar{w} is:

$$x_w = \bar{w} \cdot \bar{x} / \|\bar{w}\|$$

- If the constraint in (*) is imposed, we have:

$$x_w = \theta / \|\bar{w}\|$$

(For example: consider when $x = 2$, when $y = 3$)

- So, assuming \bar{w} and θ are constant, the projection x_w is constant and, in the case of 2D, x_w must lie along the perpendicular to the weight vector.
- Therefore, in 2D, the relation $\bar{w} \cdot \bar{x} = \theta$ defines a straight line. Generalizing to n-dimensions, the counterpart is a hyperplane.
- When x lies on the hyperplane:
 - $\bar{w} \cdot \bar{x} = \theta$ and hence $y = 1$.
- Suppose $x_w > \theta / \|\bar{w}\|$,
 - then x must lie in region A, as $\bar{w} \cdot \bar{x} > \theta$ and $\therefore y = 1$.
- If $x_w < \theta / \|\bar{w}\|$,
 - then the projection is shorter, and x must lie in region B.
- TLU is a linear classifier. If patterns cannot be separated by a hyperplane, then they cannot be classified with a single TLU.

