Neural Networks

Homework 2

Juan Jose Mendez Torrero (s3542416)
Sharif Hamed (s2562677)

April 30, 2018

1 Linear Algebra

1. The resulting coordinate system. See Figure 1.

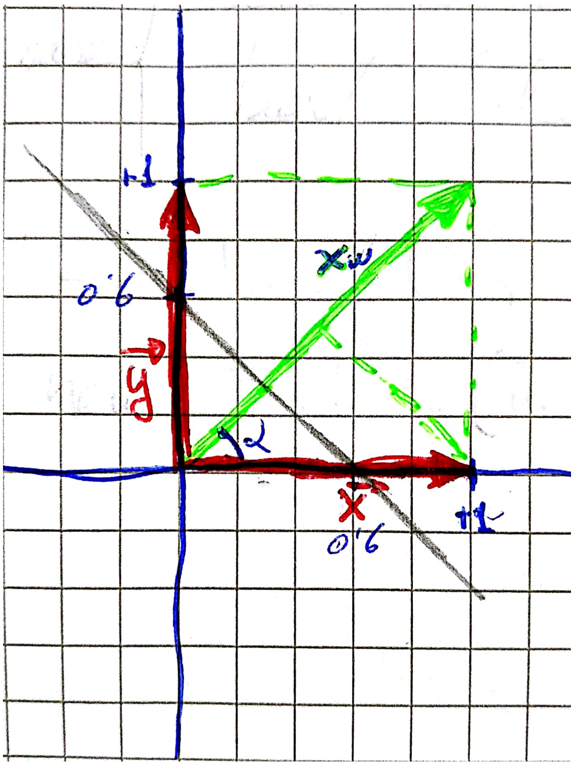

Figure 1: Coordinate system

2. We consider two vectors: $\boldsymbol{x} = \begin{vmatrix} 1 \\ 0 \end{vmatrix}$ and $\boldsymbol{y} = \begin{vmatrix} 0 \\ 1 \end{vmatrix}$

3. The length of a vector algebraically is defined as: $||x|| = \sqrt{x_1^2 + x_2^2}$. Hence, the length of $\boldsymbol{x}$ is $\sqrt{1^2 + 0^2} = 1$.

4. The inner product of two vector is defined as: $\boldsymbol{x} \cdot \boldsymbol{y} = \Sigma x_i y_i$. Where $i = 1, 2$, because the two vectors belongs to $\Re^2$. That means, that the inner product between $\boldsymbol{x}$ and $\boldsymbol{y}$ is $(1 \cdot 0) + (0 \cdot 1) = 0$.

Note that if the inner product is equal to zero, the two vectors are orthogonal between them.

5. As we can see in Figure 1, both vectors are orthogonal, so that means that our last affirmation was correct.

6. See Figure 1 for the result.

7. See Figure 1 for the result.

8. In Figure 1, we can see that the angle created with $\boldsymbol{x}_w$ and $\boldsymbol{x}$ is 45 degrees. Hence $\cos(45) = 0.7071$. Furthermore, we can determine:

$$\boldsymbol{x}_w = \frac{\boldsymbol{x} \cdot \boldsymbol{w}}{||\boldsymbol{w}||} = \frac{(1 \cdot 1) + (1 \cdot 0)}{\sqrt{1^2 + 1^2}} = \frac{1}{\sqrt{2}} = 0.7071$$

With this, we have demonstrated $\cos(\phi) = \frac{\boldsymbol{x}_w}{||\boldsymbol{x}||}$

9.

$$\boldsymbol{x} \cdot \boldsymbol{w} = ||\boldsymbol{x}|| \cdot ||\boldsymbol{w}|| \cdot \cos(\phi)$$

$$\cos(\phi) = \frac{\boldsymbol{x}_w}{||\boldsymbol{x}||}$$

$$\boldsymbol{x} \cdot \boldsymbol{w} = ||\boldsymbol{x}|| \cdot ||\boldsymbol{w}|| \cdot \frac{\boldsymbol{x}_w}{||\boldsymbol{x}||}$$

$$\boldsymbol{x} \cdot \boldsymbol{w} = ||\boldsymbol{w}|| \cdot \boldsymbol{x}_w$$

10. $||\boldsymbol{w}|| = \sqrt{1^2 + 1^2} = \sqrt{2}$. If we want $(\boldsymbol{a} \cdot \boldsymbol{w}) - \theta \geq 0$ satisfied. Then, $a_1 \cdot w_1 + a_2 \cdot w_2 \geq 0.6$. If we substitute the vector $\boldsymbol{w}$, then, we will have $a_1 + a_2 \geq 0.6$. From here, we can deduce $\boldsymbol{a} = (0.6 - a_2, a_2)$. This will satisfied the first equation for any positive value of a_2.

11. See Figure 1 for the result.

2 A TLU on paper

1. Figure 2 represent a TLU on paper.

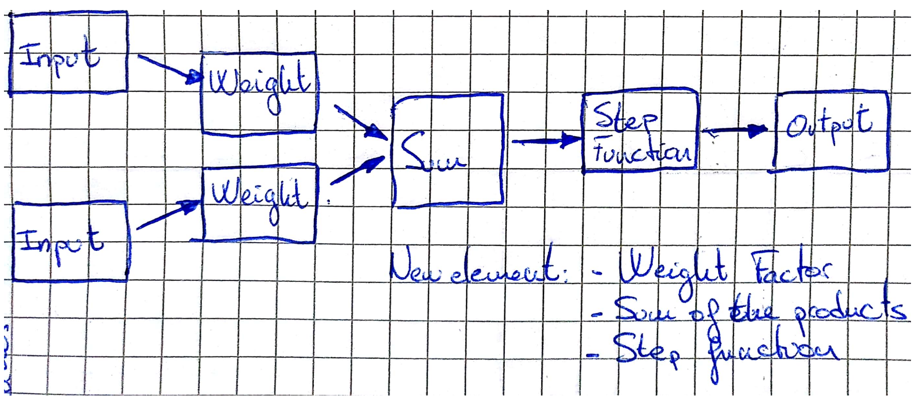

Figure 2: TLU on paper

2. As we know, if we take as inputs the dendrites of a neuron, it weight factor represent the inhibitory or excitatory of the synapse. That means that the weight can take positive or negative values.
   About the sum of the product between the inputs and their weight, we can compare it with the neuron's soma.
   Finally, about the step function, we can say that it represent the axon of a neuron.
3. Hyperpolarization could occur when an ion channel of the membrane, opens or closes. This, alters the ability, of particular types of ion, to enter or exit the cell.

3 Perceptron rule

1.

4 Delta rule

1.

5 Logistic model

1.

6 Essay questions

1. As we know, the mean-squared error is defined as follows:

$$e^p = \frac{1}{N}\Sigma(e^p)$$

   Where $e^p = \frac{1}{2}(t^p - a^p)^2$.
   The attributes that make the mean-squared error a good error function are:

   (a) The squared factor. This is really important because without it, the values of the mean-squared error could take a negative value. In order to eliminate that value, we add the squared function of e^p.

   (b) The other attribute is the activator a^p. We use this in order to prevent e^p to be a suitable candidate for gradient descent. What we do here is substitute the output with the activation variable.

   (c) The last important attribute is the factor $\frac{1}{2}$, and it has to do with the simplification of the resulting slopes or derivates.

2. First, the main difference between them is, the perceptron rule is for thresholded perceptrons, and the delta rule is for linear unit or unthresholded perceptrons. The other difference is that if the target values cannot perfectly fit, then the thresholded perceptron will be correct whenever the linear unit has the right sign. Furthermore, the error in delta rule is not restricted to having values 0, 1 or -1, as in perceptron rule. The last difference is that the delta rule can be derived for any differentiable output/activation function f, whereas the perceptron rule only works for thresholded output function.