DL for NLP - Exercise 1

Itamar Trainin - 315425967

Dec 2019

Question 1:

It is not possible to find a 2-dimensional vector w such that we get xor operation. This is because if we were to plot the 1's and 0's on a two dimensional graph: (0,0) = 0, (0,1) = 1, (1,0) = 1, (1,1)=0 we are not able to fit a straight line that divides the two binary classes.

Question 2:

a.

In this case the first layer h acts as a transformation and translation layer, and the second layer $f(x)$ acts as a classification layer. Therefore we would like to use the first layer to transform the results of the xor operation such that it would be possible to divide them into two group linearly.
Here U acts as a transformation which can be used as skew-matrix. b_2 is the translation vector. One option is to use the following transformation and translation to make the function linearly separable:

The skew can be achieved choosing $U = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, then the translation to the right is done by choosing $b_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$. Note that after the first layer

$$\begin{aligned} x = (0,0) &\rightarrow h = (0,0) \\ x = (0,1) &\rightarrow h = (0,1) \\ x = (1,0) &\rightarrow h = (0,1) \\ x = (1,1) &\rightarrow h = (1,2) \end{aligned} \tag{1}$$

Finally we want to choose w, b_1 such that it represents a linear classification boundary.That is we want:

$$\begin{aligned} sign(w \cdot x + b_1) > 0;\ h = (0,1) \\ sign(w \cdot x + b_1) < 0;\ h = (0,0), (1,2) \end{aligned} \tag{2}$$

This can be easily achieved by setting $w = \begin{pmatrix} -4 \\ 2 \end{pmatrix}$ and $b_1 = -1$. Therefore now we have

$$\begin{aligned} x = (0,0) \to h = (0,0) \to f(x) = -1 \to sign(f(x)) = -1 \\ x = (0,1) \to h = (0,1) \to f(x) = 1 \to sign(f(x)) = 1 \\ x = (1,0) \to h = (0,1) \to f(x) = 1 \to sign(f(x)) = 1 \\ x = (1,1) \to h = (1,2) \to f(x) = -1 \to sign(f(x)) = -1 \end{aligned} \tag{3}$$

which is exactly the xor function.

b.

Finding a linear classifier to this problem without the max function is not possible because we must have a non-linear transformation in-order to change the data values such that they could be pass a straight line between them.