University of Warwick
Department of Computer Science

CS331

Neural Computing

Cem Yilmaz
March 1, 2024

1 Question 1

1.1 (a)

For the function

$$f(x_1, x_2, x_3) = (x_1 \text{NIMPLY} x_2) \text{NIMPLY} x_3$$

We can construct its truth table as follows:

Table 1: $f(x_1, x_2, x_3)$ truth table

x_1	x_2	x_3	$f(x_1, x_2, x_3)$
F	F	F	F
F	F	T	F
F	T	F	F
F	T	T	F
T	F	F	T
T	F	T	F
T	T	F	F
T	T	T	F

From this table, notice that we only have a single T value, which is true if x_1 is true, x_2 is false and x_3 is false. Therefore, we can deduce that it is always false whenever x_2 or x_3 are true, as such, we can define them as inhibitory, i.e., when x_2 or x_3 is $T = 1$, the output is always $F = 0$. This leaves us a case at which f as a neuron can only fire when $x_2, x_3 = 0$, which means that now it is only dependent on x_1. There are two combinations: $x_1 = 1$ or $x_1 = 0$. In this case, we just have to define $\theta = 1$, that is, $z \geq 1$ to ensure that we fire when x_1 is $T = 1$.

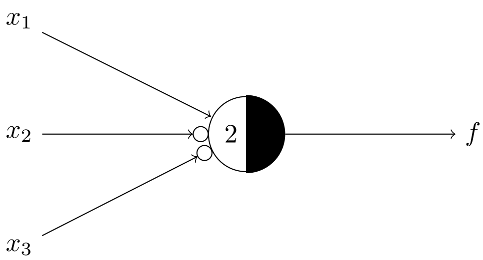

Figure 1: Rojas diagram of f

1.2 (b)

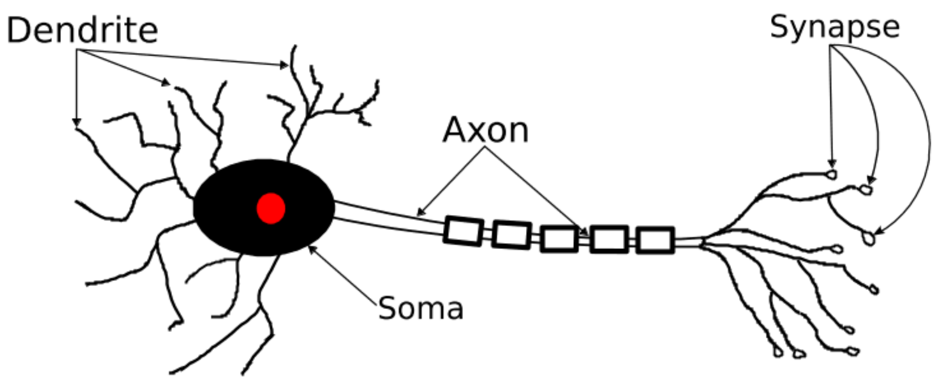

Figure 2: Biological Neuron

Two main factors that influence the speed of signal transmission along the axon:

1. Myelination - myelin acts as an insulator for the charge, creating saltatory conduction. That is, through the process of polarisation and depolarisation of Na^{+} and K^{+} ions, it is able to create action potential between each myelin which allows it to skip distances.
2. Diameter - higher diameter of axon means that there is less resistance during signal transmission as the electrical charge (ions) flow more freely.

1.3 (c)

f can be emulated by a single-layer perceptron. Consider a 3D plane with its axis defined as x_1, x_2, x_3. We notice that the function forms a discrete $2 \times 2 \times 2$ cube, for which $(1, 1, 1)$ resides on the corner. We can simply create a slanted flat 2D plane $x + y + z = 2.5$ that cuts the corner only as shown in the figure below:

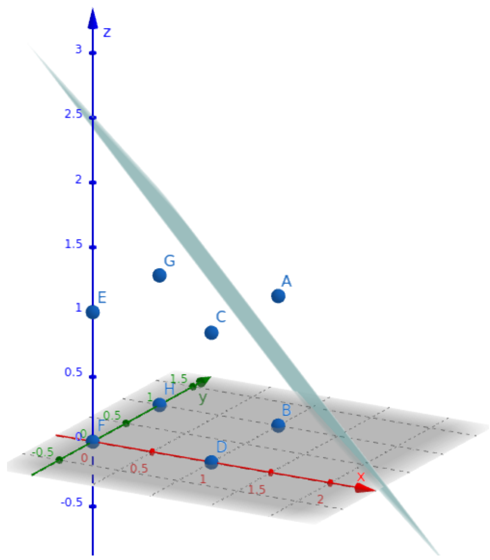

Figure 3: f mapping and separating $(1, 1, 1)$ only

We know that $(1, 1, 1)$ is the only combination for which the sum is 3, with all other sums being ≤ 2. Using this information, we can simply modify such that when we use the step function that all results ≤ 2 give $H(z_{\neg(1,1,1)}) = 0$ and $H(z_{(1,1,1)}) = 1$ one by adding -3. That is, we know that $b = -3$ if we consider $\forall w_i = 1$, $i \in \{1, 2, 3\}$ intuitvely. Therefore, the perceptron diagram for function f can be depicted as such:

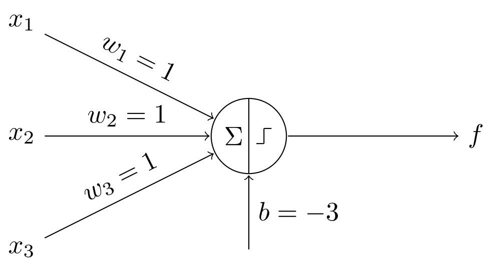

Figure 4: Rojas diagram of f

2 Question 2

2.1 (a)

For hidden layer h_1, we find the input is

$$1 \cdot -4 + -2 \cdot -2 - 1 = -1$$

Similarly, for h_2 and h_3 respectively it is

$$\begin{aligned} 1 \cdot 1 + -2 \cdot 1 + 0 = -1 \\ 1 \cdot 0 + -2 \cdot 1 + 2 = 0 \end{aligned}$$

Therefore, we find that the output of the hidden layer is, for each hidden layer,

$$\begin{aligned} \text{tanh}_1(-1) &= -0.76159415595\ldots \\ \text{tanh}_2(-1) &= -0.76159415595\ldots \\ \text{tanh}_3(0) &= 0 \end{aligned}$$

The input to the output layer, is then

$$-0.76159415595 \cdot 2 + -0.76159415595 \cdot -1 + 0 \cdot 1 + 2 = 1.23840584405$$

Therefore, the output $\hat{y}$ is then found to be

$$\sigma(1.23840584405) = 0.7753 \;\; \text{4d.p.}$$

2.2 (b)

We list all the related equations as such:

$$\begin{aligned} z_1 &= w_1 x_1 + w_4 x_2 + b_1 \\ h_1 &= \text{tanh}(z_1) \\ z_y &= h_1 w_7 + h_2 w_8 + h_3 w_9 \\ \hat{y} &= \sigma(z_y) \\ L &= \frac{1}{2}\left(y - \hat{y}\right)^2 \end{aligned}$$

Therefore for the partial derivatives we obtain

$$\begin{aligned} z_{1w_1} &= x_1 \\ h_{1z_1} &= 1 - h_1^2 \\ z_{yh_1} &= w_7 \\ \hat{y}_{z_y} &= \hat{y}\left(1 - \hat{y}\right) \\ L_{\hat{y}} &= \hat{y} - y \end{aligned}$$

We now multiply all of these values that create a single path as such:

$$x_1 \cdot (1 - h_1^2) \cdot w_7 \cdot \hat{y}(1 - \hat{y}) \cdot (\hat{y} - y)$$

Which after substitution obtains us

$$g = 1 \cdot (1 - (-0.76159415595)^2) \cdot 2 \cdot 0.77528640646(1 - 0.77528640464) \cdot (0.77528640646 - 0.7)$$

For which the above computes to

$$g = 0.01 \;\; \text{2d.p.}$$

3 Question 3

3.1 (a)

Two disadvantages to the sigmoid function in training neural networks:

1. Vanishing gradient - large or small inputs have very little change in prediction. This can result in the network refusing to learn further.
2. Computationally expensive - the present of exponential function means that the sigmoid function is computationally heavy.

We first show that the sigmoid function $\sigma(x)$ has a property of being strictly increasing and non-negative. We first prove that it is strictly non-negative:

Proof. We have that, $\forall x$ the function $e^{-x} > 0$. Using this information, we can deduce that the function $\sigma(x)$

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Is also strictly non-negative. □

We now show it is strictly increasing.

Proof. That is, by differentiating, we find

$$\sigma'(x) = \frac{e^x}{(e^x + 1)^2}$$

We know that $\forall x, e^x > 0$, therefore, we can deduce that $\forall x, \sigma'(x) > 0$. We have shown that it is strictly increasing, a property which will be useful in a later proof. □

For a function $f(x)$ to be a probability density function, the requirement is that

$$\int_{-\infty}^{+\infty} f(x)\ \mathrm{d}x = 1$$

Otherwise, we can actually split the integral such that

$$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = \int_{-\infty}^{1} f(x)\mathrm{d}x + \int_{1}^{5} f(x)\mathrm{d}x + \int_{5}^{+\infty} f(x)\mathrm{d}x$$

We now calculate

$$\int_{1}^{5} \sigma(x)\mathrm{d}x = \left[\ln\left(e^{-x} + 1\right)\right]_1^5 = 3.693\ldots$$

However, we know that the function is strictly increasing and non-negative. Therefore, we can deduce that

$$\int_{-\infty}^{1} \sigma(x)\mathrm{d}x > 0$$
$$\int_{5}^{+\infty} \sigma(x)\mathrm{d}x > 0$$

Therefore, the sum of the three integrals, denote it y, achieves us

$$y > 3.693\ldots$$

which proves that it is not a probability density function.

3.2 (b)

We first show similarity using Jaccard node-pair similarity. We have that

$$\text{sim}_J\,(5,6) = \frac{|I(5) \cap I(6)|}{|I(5) \cup I(6)|}$$

We find that

$$I(5) = \{2,3\}$$
$$I(6) = \{1,2,3\}$$

Using the graph of network G. Then, substituting we find

$$\begin{aligned}\text{sim}_J\,(5,6) &= \frac{|\{1,2\}|}{|\{1,2,3\}|} \\ &= \frac{2}{3} \\ &= 0.67 \;\; \text{2d.p.}\end{aligned}$$

We now compute using SimRank. We have that $I(5) \neq I(6)$, therefore

$$\text{sim}_R(5,6) = \frac{0.6}{|I(5)||I(6)|} \sum_{x \in I(5)} \sum_{y \in I(6)} \text{sim}_R(x,y)$$

We obtain

$$\frac{0.6}{|I(5)||I(6)|} \left(\text{sim}_R(3,1) + \text{sim}_R(3,2) + \text{sim}_R(3,3) + \text{sim}_R(2,1) + \text{sim}_R(2,2) + \text{sim}_R(2,3)\right)$$

Knowing $\text{sim}_R(z,z) = 1$ we can simplify

$$\frac{0.6}{|I(5)||I(6)|} \left(\text{sim}_R(3,1) + \text{sim}_R(3,2) + \text{sim}_R(2,1) + \text{sim}_R(2,3)\right)$$

We can further simplify by knowing the fact that $I(1) = \varnothing$ and that $\text{sim}_R(a,b) = \text{sim}_R(b,a)$, therefore from definition of SimRank we have

$$\frac{0.6}{|I(5)||I(6)|} \left(2 + 2 \cdot \text{sim}_R(2,3)\right)$$

We now find SimRank of $(2,3)$ as follows:

$$\text{sim}_R(2,3) = \frac{0.6}{|I(2)||I(3)|} \left(\text{sim}_R(1,1)\right) = 0.6$$

Substituting to our original equation, then, we obtain

$$\begin{aligned}\frac{0.6}{|I(5)||I(6)|} (2 + 1.2) &= 0.32 \\ &= 0.32 \;\; \text{2d.p.}\end{aligned}$$

3.3 (c)