HomeWork 1

João André Roque Costa – 99088

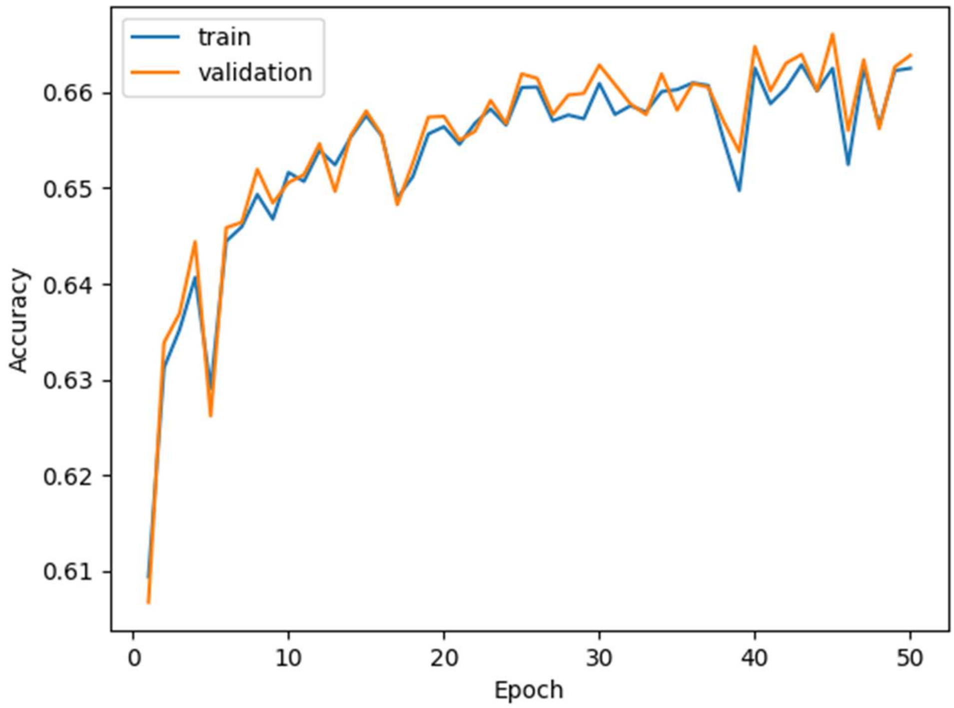
Tomás Augusto Vilhena de Oliveira – 90781

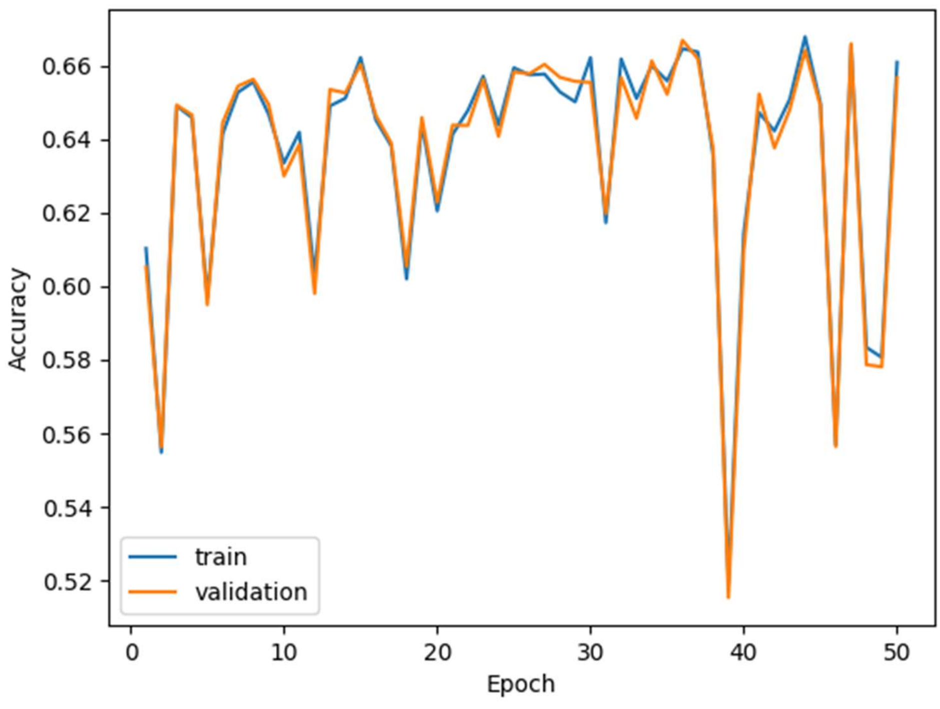
Question 1

Q1.1

The final accuracy was 0.3422, while the train accuracy was 0.4654 and the validation accuracy of 0.4610.

The final accuracy was 0.3422, while the train accuracy was 0.4654 and the validation accuracy of 0.4610.





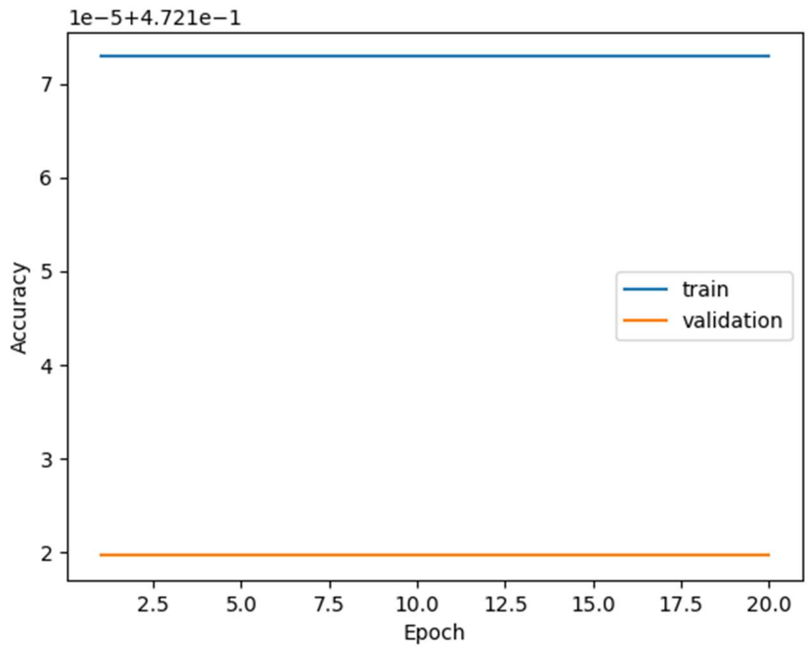
Q1.2a

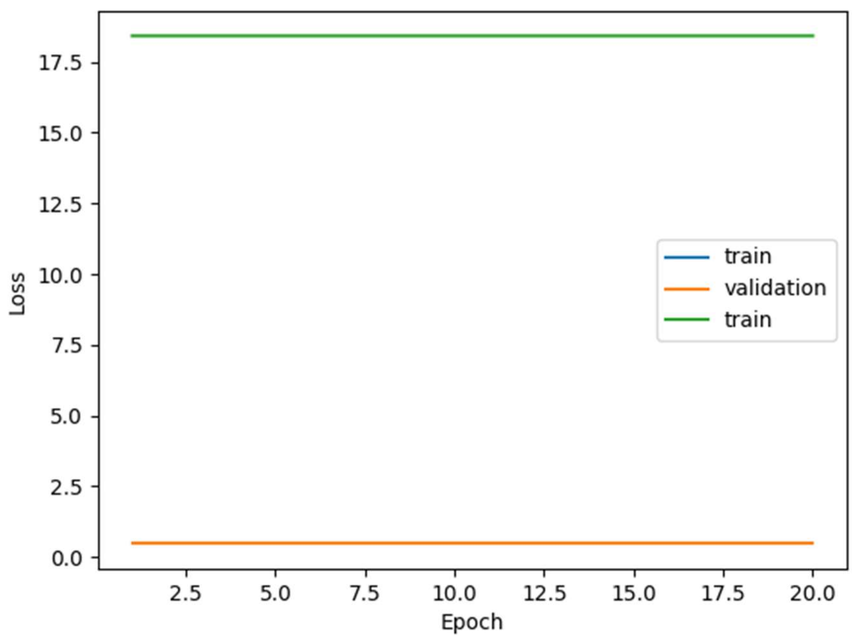
This statement is true. In terms of expressiveness, a logistic regression model is a linear model while a multi-layer perceptron using relu isn't. This means that logistic regression can at most create a linear frontier, while a mlp using relu is capable of defining nonlinear frontiers.

On terms of ease of training, logistic regression wins due to the fact that is convex, meaning that there is only one global solution, and the solution is able to converge into it. Meanwhile, in a MLP using relu, the optimization is not convex.

Q1.2b

The final accuracy was 0.4726, while the train accuracy was 0.4722 and the validation accuracy of 0.4721.

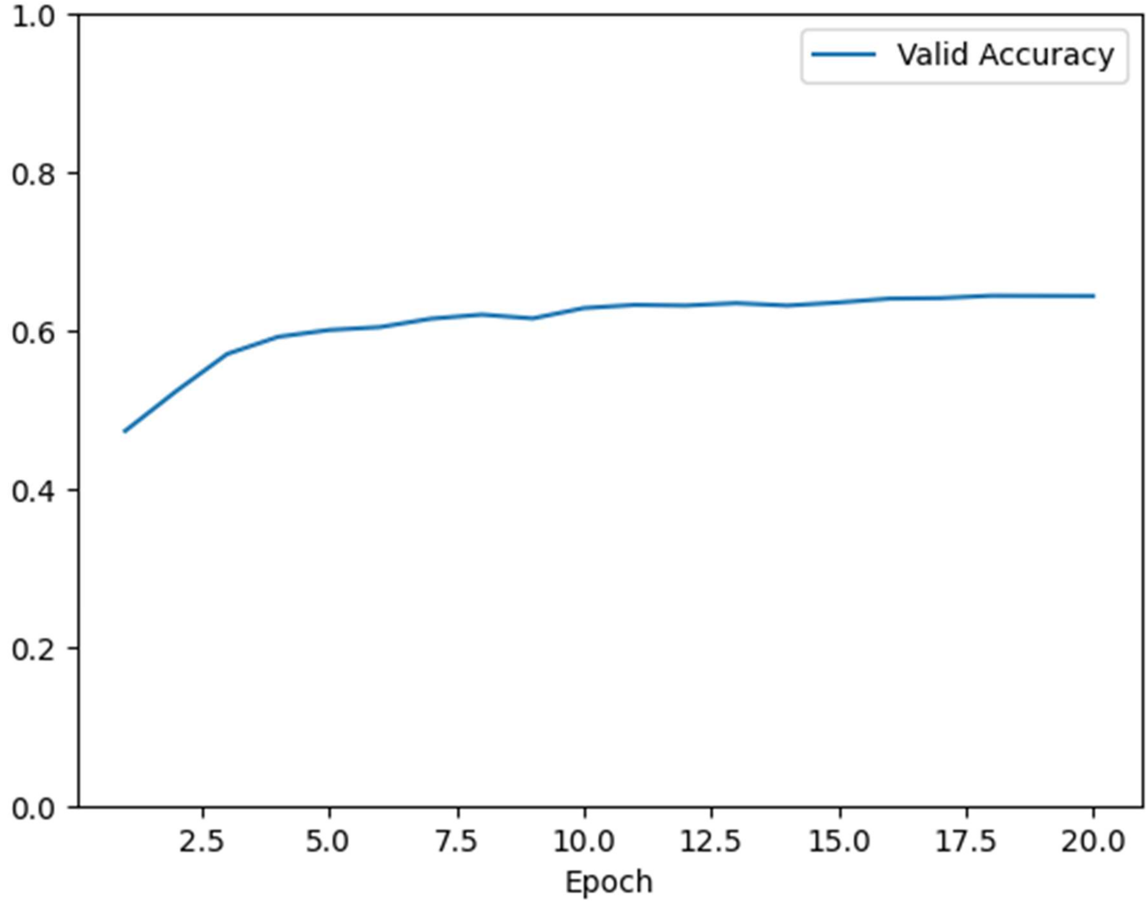


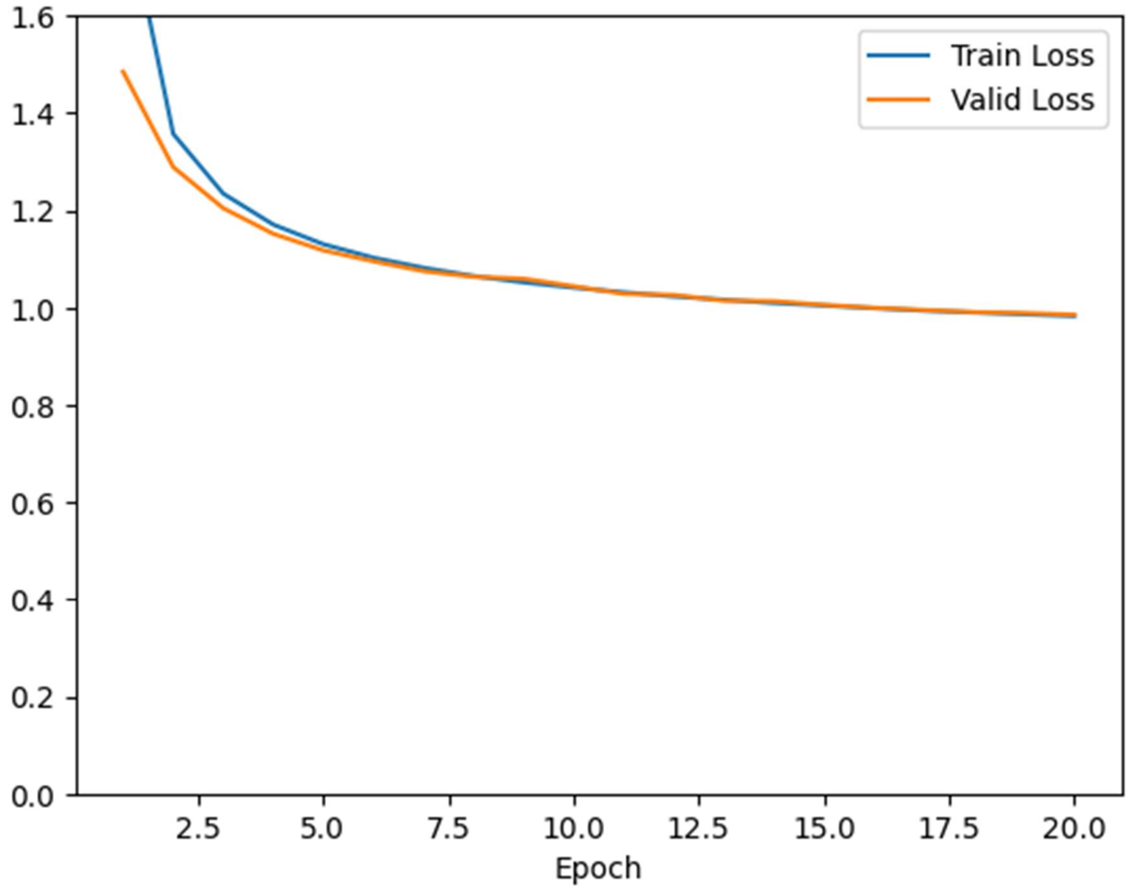


Question 2

Q2.1

The best configuration is with a learning rate equal to 0.01 with a final test accuracy of 0.6030, a validation accuracy of 0.6436 and a training loss of 0.9827.

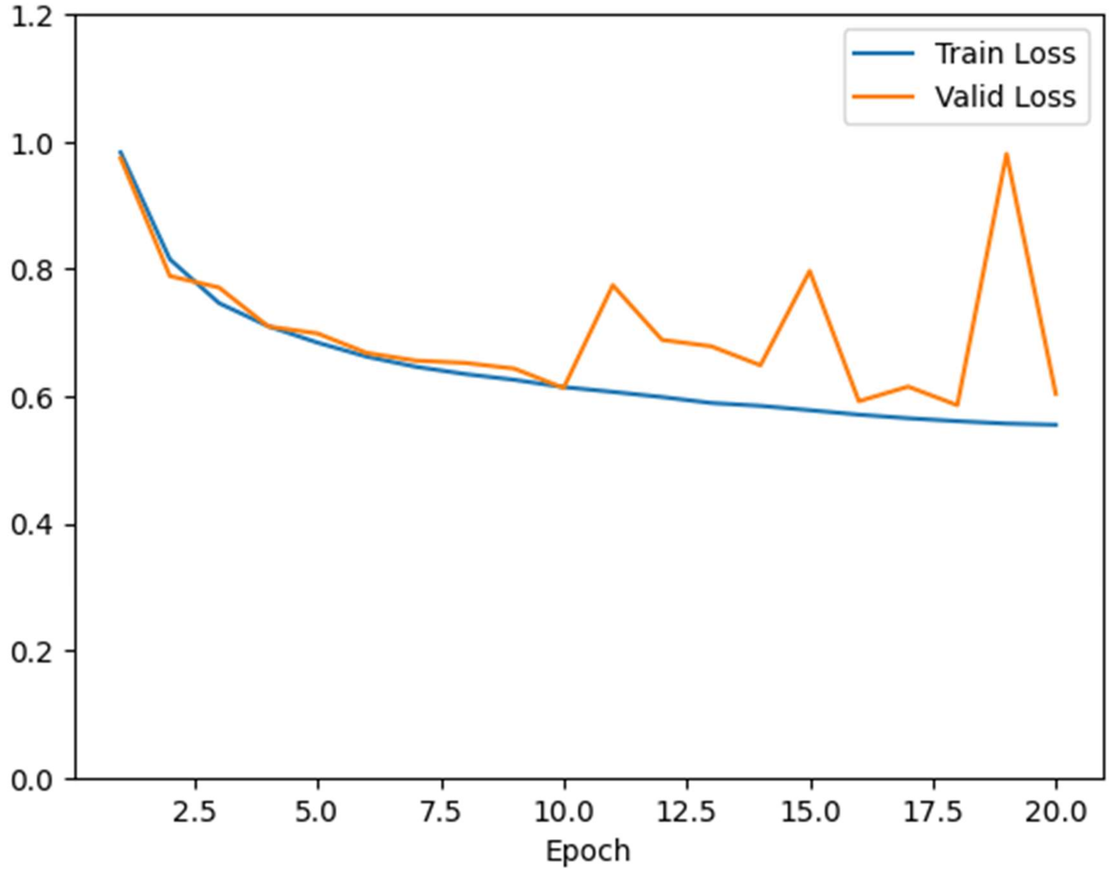


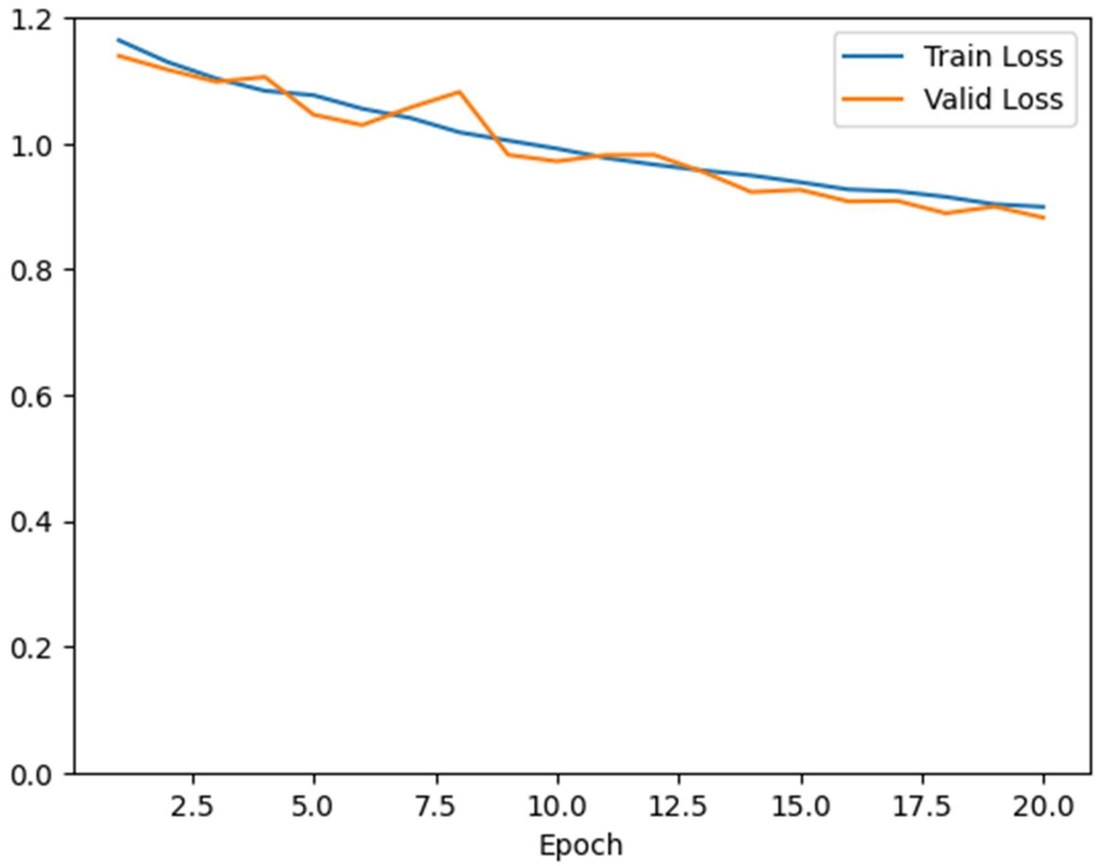


Q2.2a

Batch size of 16 took 2.219s sys time and got a final test accuracy of 0.7618, while a batch size of 1024 took 0.778s sys time and got a final test accuracy of 0.6975.

Meaning that a batch size of 16 got a better performance but took longer to train. Havin a training loss of 0.5549 and a valid accuracy of 0.7960.

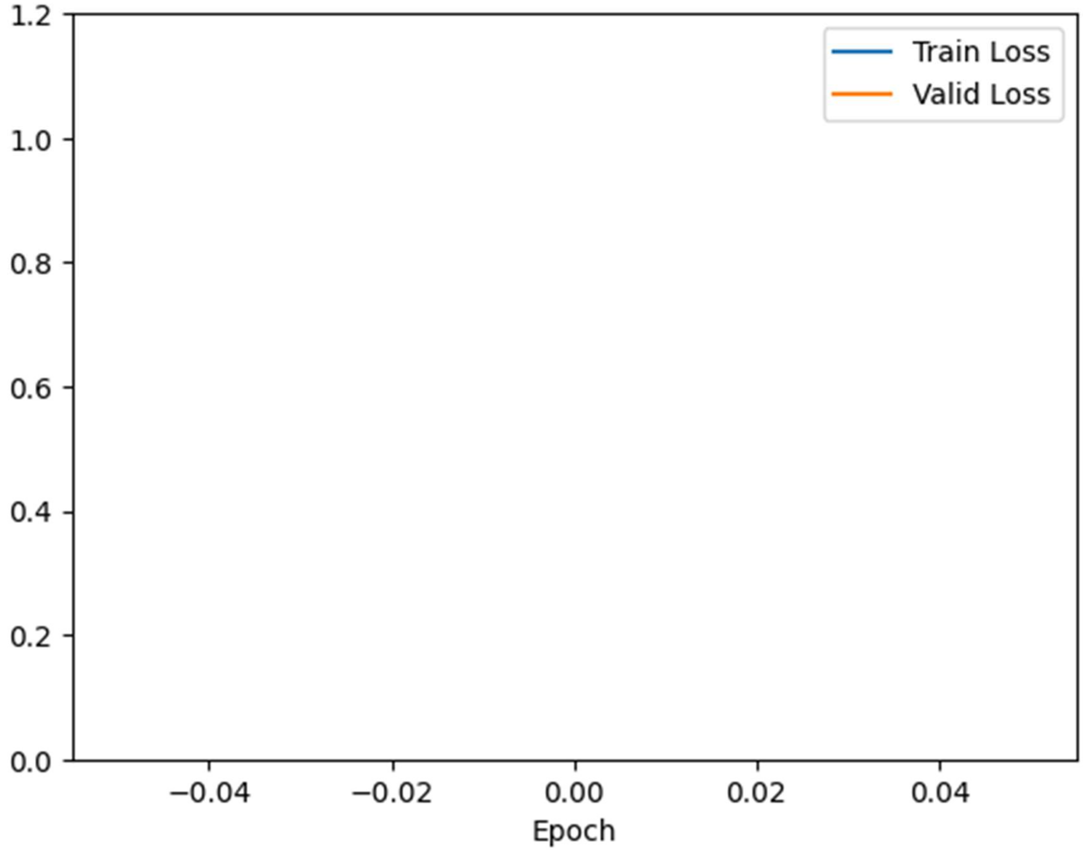


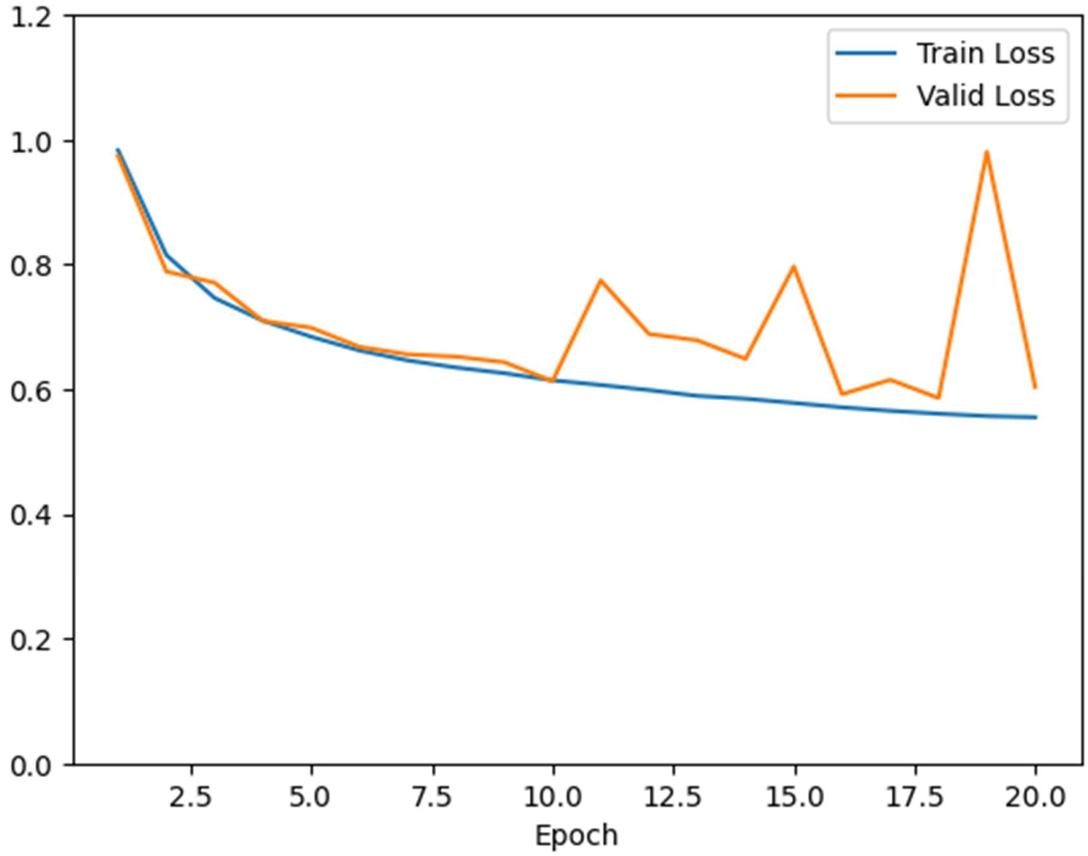


Q2.2b

The learning rate that got the best test accuracy was a learning rate of 0.1 with a test accuracy of 0.7618.

The best validation accuracy was 0.7960 for a learning rate of 0.1 and the worst one was 0.3435 for a learning rate of 1. Having the following graphs:

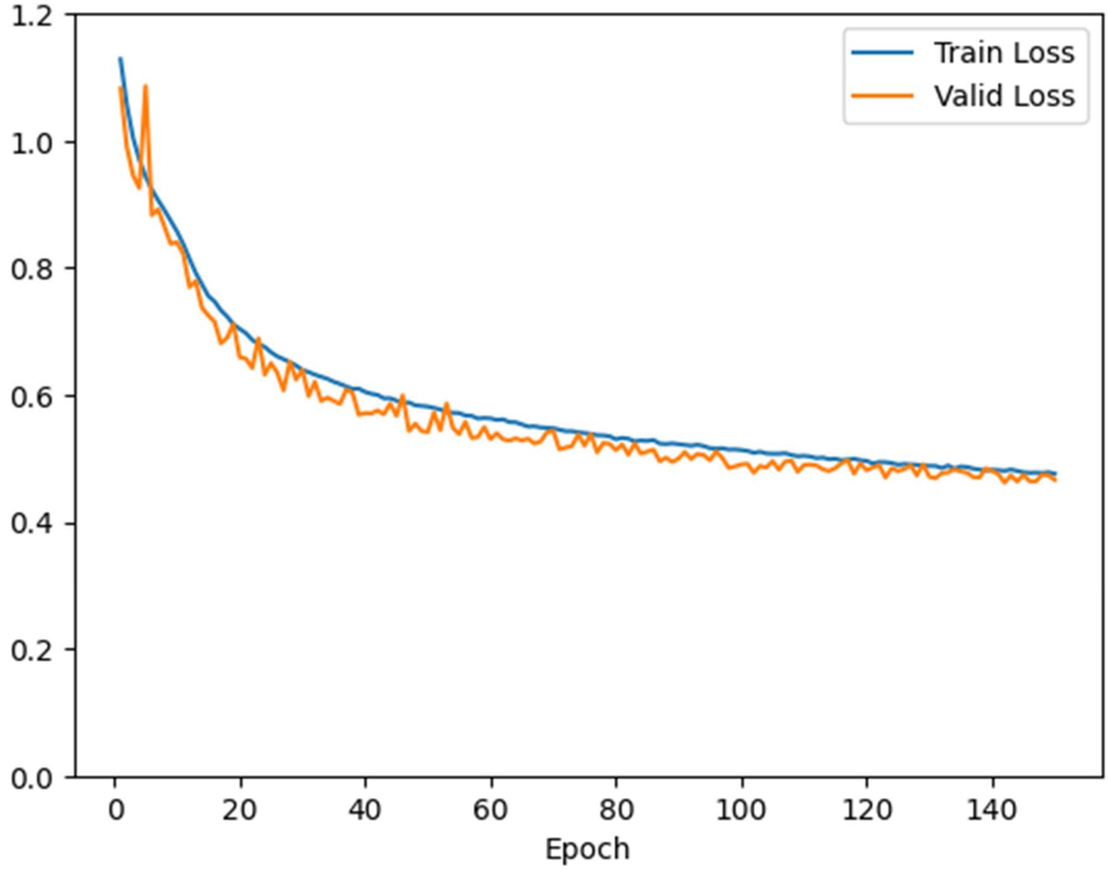


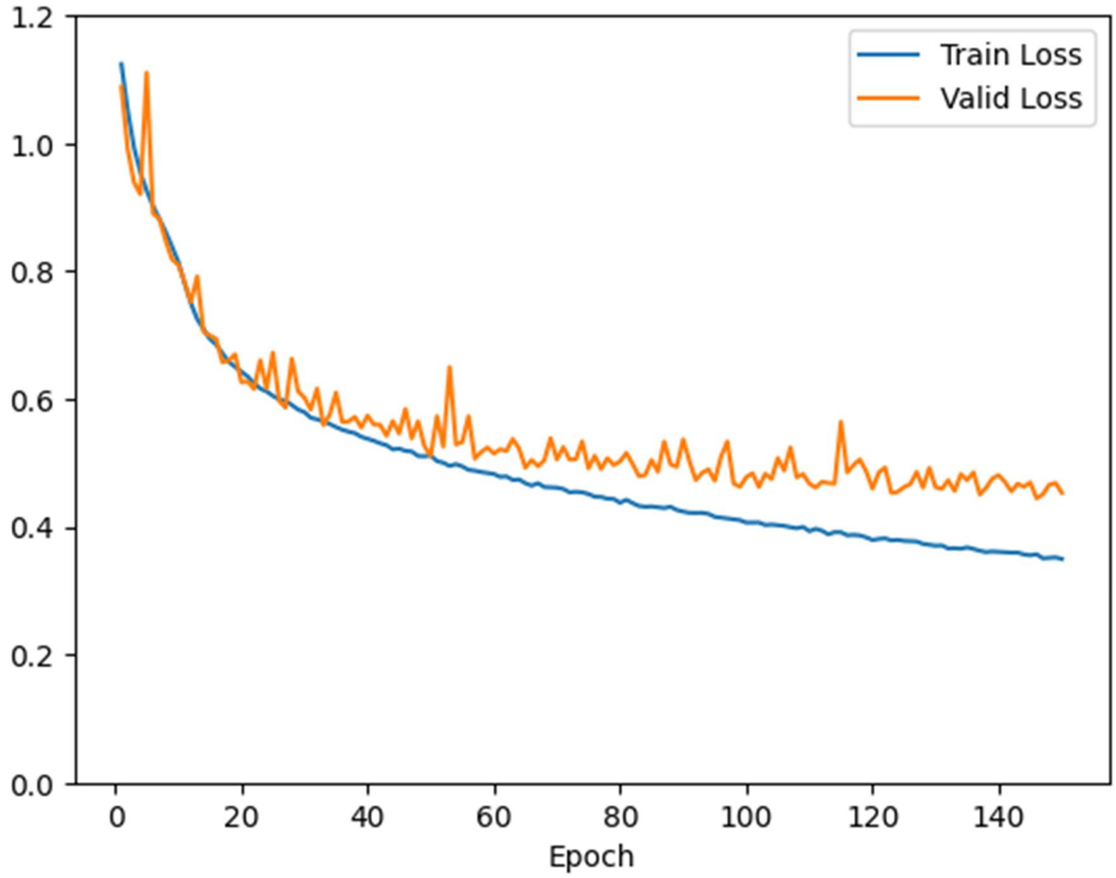


Q2.2c

Running a batch size of 256 with 150 epochs resulted in a training loss of 0.3164, a validation accuracy of 0.8428 and a test accuracy of 0.7713. Since there doesn't seem to be a decrease in validation accuracy in favor of a increase in test accuracy, there doesn't seem to be overfitting.

Setting L2 regularization to 0.0001 yelded the best validation accuracy of 0.8444 and having a dropout probability of 0.2 got the worst validation accuracy of 0.8342. Having the graphs:



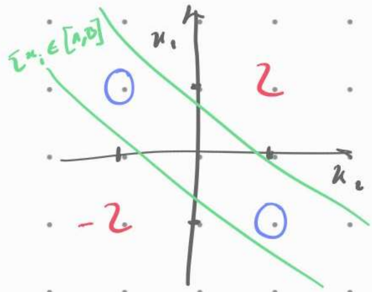


Question 3

a)

$x_i \in \{-1, 1\}, \quad 0 < i < D,$

when we consider $D = 2$:

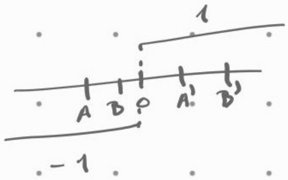


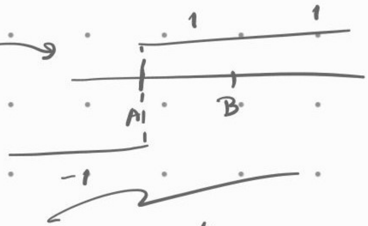
the fuction has a XOR like behavior, which is a known non linearly seperable function / single layered perceptron.

b) $h(x) = \text{sign}(w\, x + b)$

1st node : $\underbrace{\begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}_{(1\times D)} \begin{bmatrix} x_1 \\ \vdots \\ x_D \end{bmatrix}_{(D\times 1)}}_{\sum_{i=1}^{D} 1\times x_i \Rightarrow \sum x} + \underbrace{\overset{1\times 1}{0}}_{1 \text{ node} \Rightarrow 1 \text{ dim.}}$

$\text{Sign}(\sum x) \Longrightarrow$



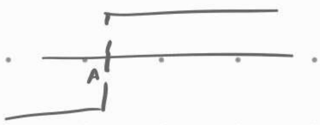


$b \simeq A$

$\underbrace{\text{sign}(w\, x + A)}$

true : $\sum x + A \geq 0$

$\sum x \geq +A$



$\underbrace{}_{\sum x \geq A}$

1st node : $w_1 = [1\ 1 \dots]_{(1\times D)}$ $\quad b_1 = -A$

2nd node :

$\sum x \leq B$

$0 \leq B - \sum x$

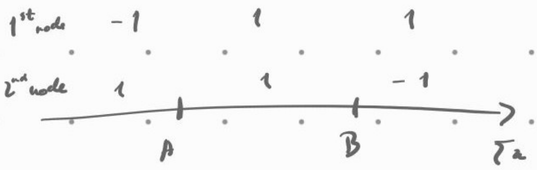
$\underline{\underline{-\sum x + B \geq 0}}$

$\text{Sign}(-\sum x + B) = \begin{cases} 1, & \sum x \leq B \\ -1, & \sum x > B \end{cases}$

2nd node : $w_2 = [-1\ -1 \dots -1]_{(1\times D)}$ $\quad b_2 = B$

Hidden layer : $w^{(1)} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ -1 & -1 & \dots & -1 \end{bmatrix}_{(2\times D)}$ $\qquad b^{(1)} = \begin{bmatrix} -A \\ B \end{bmatrix}$

Output layer :



$$\begin{cases} \begin{bmatrix} -1 \\ 1 \end{bmatrix}, & \sum x < A < B \\ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, & A \leq \sum x \leq B \\ \begin{bmatrix} 1 \\ -1 \end{bmatrix}, & B < \sum x \end{cases}$$

$\Downarrow$

$\begin{bmatrix} -1 \\ -1 \end{bmatrix} \Rightarrow \underline{\sum x < A \ \wedge \ B < \sum x}$

$\Leftrightarrow B < \sum x < A$

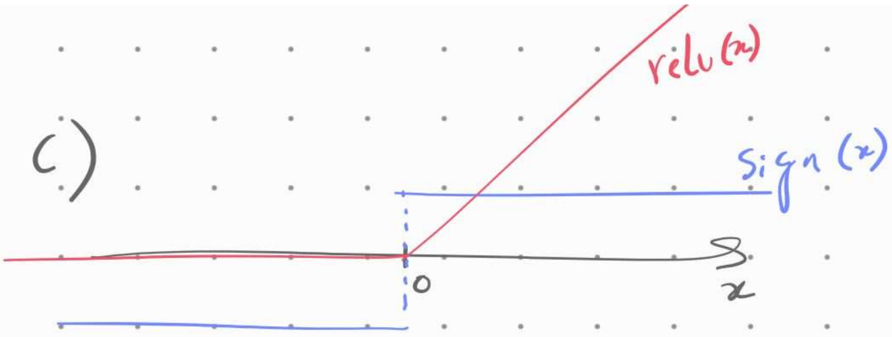
$\Leftrightarrow \underline{B < A}$

Impossible, since $-D \leq A \leq B \leq D$

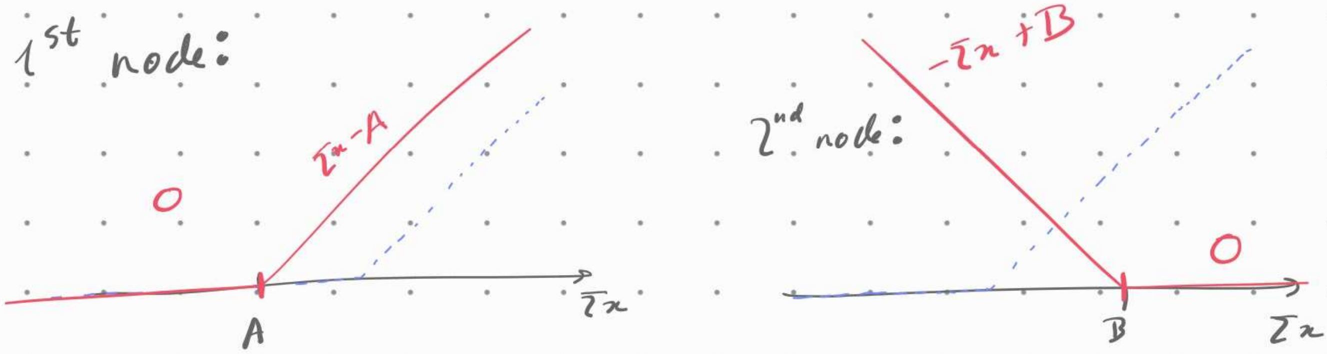
$$o(x) = w\, h^{(x)} + b \in \{-1, 1\}$$

$$\begin{cases} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \cancel{+0}\ -1 = 0 - 1 = -1 \\ \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cancel{+0}\ -1 = 2 - 1 = 1 \\ \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \cancel{+0}\ -1 = 0 - 1 = -1 \end{cases}$$

c)



Since both sign and relu
can separate classes when $x>0$ and $x<0$:



both $w^{(1)}$ and $b^{(1)}$ from the Hidden layer are the same as in b) since sign and relu separate classes on $x=0$.

$\underbrace{\Sigma x - A \geq 0}_{\Sigma x \geq 0} \Rightarrow \Sigma x - A \mid 0$

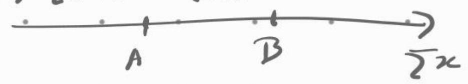
$$\begin{cases} \Sigma x - A, & \Sigma x \geq A \\ 0, & \Sigma x < A \end{cases}$$

$-\Sigma x + B \geq 0 \Rightarrow -\Sigma x + B \mid 0$

$B \geq \Sigma x$

$$\begin{cases} -\Sigma x + B, & \Sigma x \leq B \\ 0, & \Sigma x > B \end{cases}$$

$$A: \quad 0 \qquad \Sigma x - A \qquad \Sigma x - A$$
$$B: \quad -\Sigma x + B \qquad -\Sigma x + B \qquad 0$$



$$\Rightarrow \begin{cases} w^{(2)} \begin{bmatrix} \Sigma x - A \\ 0 \end{bmatrix} + b \longrightarrow -1 \\ w^{(2)} \begin{bmatrix} \Sigma x - A \\ -\Sigma x + B \end{bmatrix} + b \longrightarrow 1 \\ w^{(2)} \begin{bmatrix} 0 \\ -\Sigma x + B \end{bmatrix} + b \longrightarrow -1 \end{cases}$$

$$\begin{cases} w_1^{(2)} (\Sigma x - A) + w_2^{(2)} (0) + b = -1 \\ w_1^{(2)} (\Sigma x - A) + w_2^{(2)} (-\Sigma x + B) + b = 1 \\ w_1^{(2)} (0) + w_2^{(2)} (-\Sigma x + B) + b = 0 \end{cases} \qquad \begin{cases} w_1^{(2)} = \dfrac{-1-b}{\Sigma x - A} \\ \text{———} \\ w_2^{2} = \dfrac{-1-b}{-\Sigma x + B} \end{cases}$$

$$\begin{cases} \text{———} \\ (-1-b) + (-1-b) + b = 1 \\ \text{———} \end{cases} \qquad \begin{cases} w_1^{(2)} = \dfrac{2}{\Sigma x - A} \\ b = -3 \\ w_2^{(2)} = \dfrac{2}{-\Sigma x + B} \end{cases}$$

$$W^{(1)} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ -1 & -1 & \cdots & 1 \end{bmatrix} \qquad b^{(1)} = \begin{bmatrix} -A \\ B \end{bmatrix}$$

$$W^{(2)} = \begin{bmatrix} \dfrac{2}{\Sigma x - A} \\ \dfrac{2}{-\Sigma x + B} \end{bmatrix} \qquad b^{(2)} = -3$$

While João Costa came up with some initial resolutions to the first and second part of question 1, its quality didn't fit the standard by the same and so Tomás Oliveira restarted the process and solved both question 1 and 2 from the start. The whole question 3 was solved by João Costa.