Deep Learning

Homework 1

Group 71 members:

- Luis Jose De Macedo Guevara (95621)

- Vincent Jakl (108529)

Contributions of each member

- Luis Jose De Macedo Guevara (95621):

- Vincent Jakl (108529):

Question 1

1. a)

The single perceptron was not the best choice for this task. As can be seen in the figure below, the perceptron was not able to fit to the high dimensional data also with no sign of improvement. With the 20 epochs, it ended up with a 0.3422 test accuracy which is slightly higher than 0.25 that would happen with random selection for 4 classes. Also the learning rate was left at the value of 1 so this might cause the problem as well.
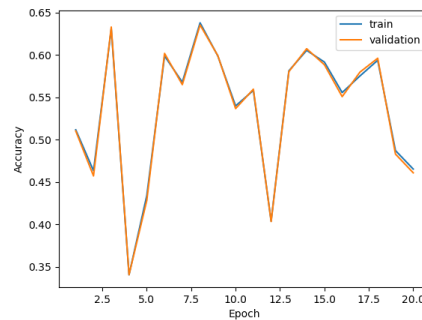

Figure 1: Single perceptron

1. b)

The logistic regression performed better than the single perceptron. As can be seen in the figure below, the logistic regression was able to fit to the data better. There was however a big difference between the two learning rates. Where the model with the learning rate of 0.01 jumped faster to the higher accuracy, it did not improve on a regular basis. The 0.001 lr model was a bit slower to get up to higher accuracy, but it was able to consistently improve. In the end the 0.01 lr model ended up with 0.5784 test accuracy and the 0.001 lr model ended up with 0.5936 test accuracy. Running these models for more epochs would possibly improve the accuracy since the trainig was still showing signs of improvement even with the evaluation metric.

2. a)

2. b)
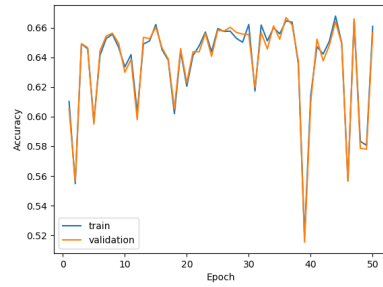
(a) lr 0.01
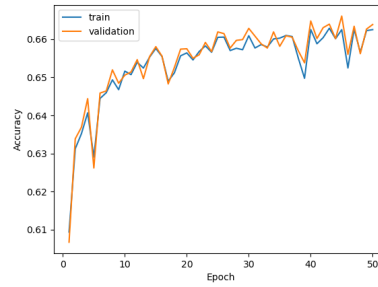
(b) lr 0.001

Figure 2: Logistic regression

Question 2

1.

2. a)

2. b)

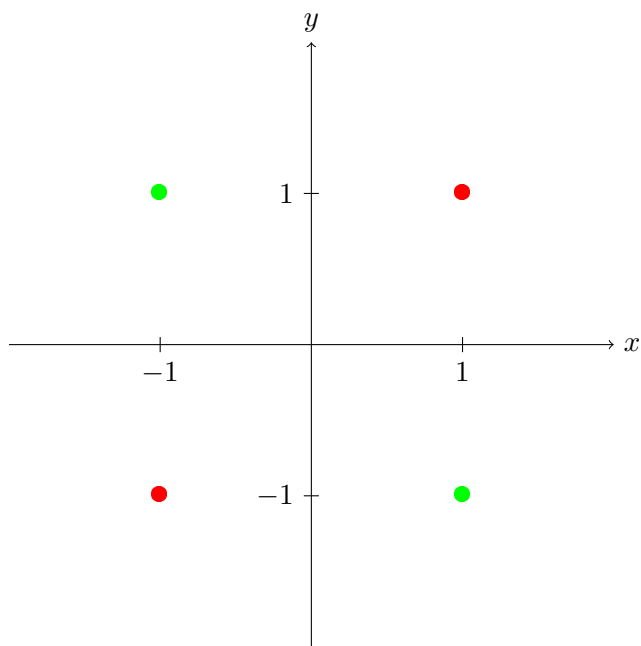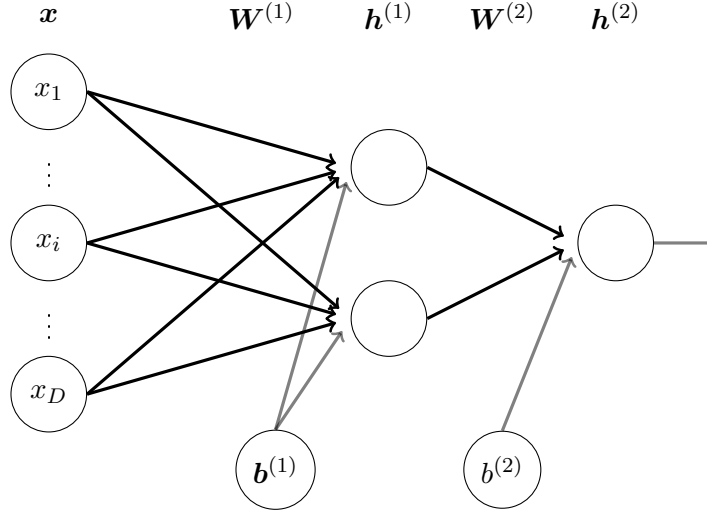**2. c)**

Question 3

1.

a)

Let

- $D = 2$

- $A = B = 0$

- $\boldsymbol{x}^{(1)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\boldsymbol{x}^{(2)} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, $\boldsymbol{x}^{(3)} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\boldsymbol{x}^{(4)} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$,

We have $f(\boldsymbol{x}^{(1)}) = -1$, $f(\boldsymbol{x}^{(2)}) = +1$, $f(\boldsymbol{x}^{(3)}) = +1$ and $f(\boldsymbol{x}^{(4)}) = -1$

Which we can see is not linearly separable and therefore a perceptron cannot learn a separating hyperplane.
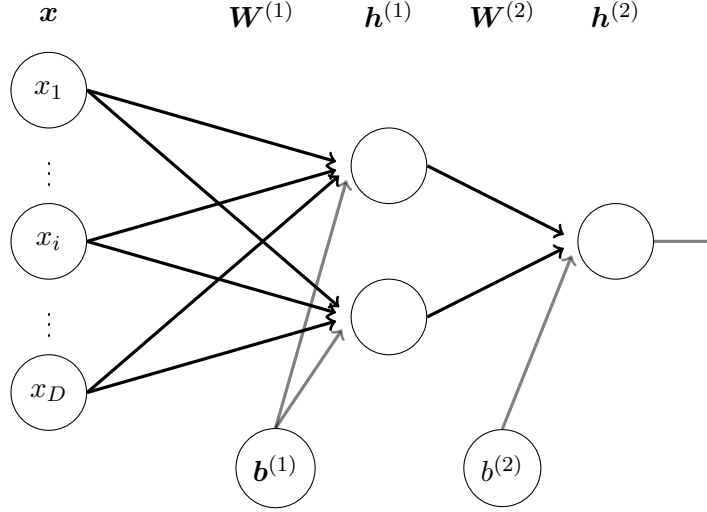
b)

$$\boldsymbol{W}^{(1)} = \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ -1 & -1 & \cdots & -1 \end{bmatrix}}_{2 \times D}, \qquad \boldsymbol{b}^{(1)} = \begin{bmatrix} -A \\ B \end{bmatrix}$$

$$\boldsymbol{W}^{(2)} = \begin{bmatrix} 1 & 1 \end{bmatrix}, \qquad b^{(2)} = -1$$

We have for $\boldsymbol{x}$:

$$\boldsymbol{h}^{(2)} = \text{sign}\left(\boldsymbol{W}^{(2)}\boldsymbol{h}^{(1)} + b^{(2)}\right)$$

$$= \text{sign}\left(\boldsymbol{W}^{(2)}\left(\text{sign}\left(\boldsymbol{W}^{(1)}\boldsymbol{x} + \boldsymbol{b}^{(1)}\right)\right) + b^{(2)}\right)$$

$$= \text{sign}\left(\begin{bmatrix} 1 & 1 \end{bmatrix}\left(\text{sign}\left(\begin{bmatrix} 1 & 1 & \cdots & 1 \\ -1 & -1 & \cdots & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix} + \begin{bmatrix} -A \\ B \end{bmatrix}\right)\right) - 1\right)$$

$$= \text{sign}\left(\begin{bmatrix} 1 & 1 \end{bmatrix}\begin{bmatrix} \text{sign}\left(\sum_{i=1}^{D} x_i - A\right) \\ \text{sign}\left(B - \sum_{i=1}^{D} x_i\right) \end{bmatrix} - 1\right)$$

$$= \text{sign}\left(\text{sign}\left(\sum_{i=1}^{D} x_i - A\right) + \text{sign}\left(B - \sum_{i=1}^{D} x_i\right) - 1\right)$$

c)

$$\boldsymbol{W}^{(1)} = \underbrace{\begin{bmatrix} -1 & -1 & \cdots & -1 \\ 1 & 1 & \cdots & 1 \end{bmatrix}}_{2 \times D}, \qquad \boldsymbol{b}^{(1)} = \begin{bmatrix} A \\ -B \end{bmatrix}$$

$$\boldsymbol{W}^{(2)} = \begin{bmatrix} -1 & -1 \end{bmatrix}, \qquad b^{(2)} = 0$$

We have for $\boldsymbol{x}$:

$$\boldsymbol{h}^{(2)} = \mathrm{sign}\left(\boldsymbol{W}^{(2)} \boldsymbol{h}^{(1)} + b^{(2)} \right)$$

$$= \mathrm{sign}\left(\boldsymbol{W}^{(2)} \left(\mathrm{ReLU}\left(\boldsymbol{W}^{(1)} \boldsymbol{x} + \boldsymbol{b}^{(1)} \right) \right) + b^{(2)} \right)$$

$$= \mathrm{sign}\left(\begin{bmatrix} -1 & -1 \end{bmatrix} \left(\mathrm{ReLU}\left(\begin{bmatrix} -1 & -1 & \cdots & -1 \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix} + \begin{bmatrix} A \\ -B \end{bmatrix} \right) \right) \right)$$

$$= \mathrm{sign}\left(\begin{bmatrix} -1 & -1 \end{bmatrix} \begin{bmatrix} \mathrm{ReLU}\left(A - \sum_{i=1}^{D} x_i \right) \\ \mathrm{ReLU}\left(\sum_{i=1}^{D} x_i - B \right) \end{bmatrix} \right)$$

$$= \mathrm{sign}\left(-\mathrm{ReLU}\left(A - \sum_{i=1}^{D} x_i \right) - \mathrm{ReLU}\left(\sum_{i=1}^{D} x_i - B \right) \right)$$