Deep learning HK2 Participants: 1. Hanzhi Zhuang 2. Rui Song 3. Tengyunhao Yang.

1. Pen and Paper MLP

$$z = w^T x + b. \quad H = g(z)$$

$$z^{(1)} = w^{(1)T} x^{(1)} + b^{(1)}$$

$$= \begin{pmatrix} -2 & -3 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 3 \\ 2 & 4 \\ 3 & 5 \end{pmatrix} + \begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} -7 & -13 \\ 4 & 8 \end{pmatrix} + \begin{pmatrix} 3 & 3 \\ 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} -4 & -10 \\ 4 & 8 \end{pmatrix}$$

$$X^{(2)} = H(1) = g^{(1)}(z^{(1)}) = \begin{pmatrix} 0 & 0 \\ 4 & 8 \end{pmatrix}$$

$$z^{(2)} = w^{(2)T} X^{(2)} + b^{(2)}$$

$$= (-1\ 1) \begin{pmatrix} 0 & 0 \\ 4 & 8 \end{pmatrix} + (-3)$$

$$= (4\ \ 8) + (-3\ -3)$$

$$= (1\ \ 5)$$

$$H(2) = \left(\frac{1}{1+e^{-1}} \quad \frac{1}{1+e^{-5}} \right) \approx (0.731 \quad 0.993)$$

$$\text{CrossEntropyError}_{(x_1, x_2)} = L(\hat{y}, y) = -y \log \hat{y} - (1-y) \log(1-\hat{y}) \approx (0.310 \quad 0.007)$$

$$L(w) = -\frac{1}{N} \sum_{n=1}^{N} CEE_{(x_n)} = (0.310 + 0.007)/2 \approx 0.159$$

3. Experiments.

1) the best accuracy from Logistic Regression in this case is 75%.

Because the XOR operation is not perfectly linear seperable, so logistic regression will not be able to achieve 100% accuracy. as can be seen in the graph below.