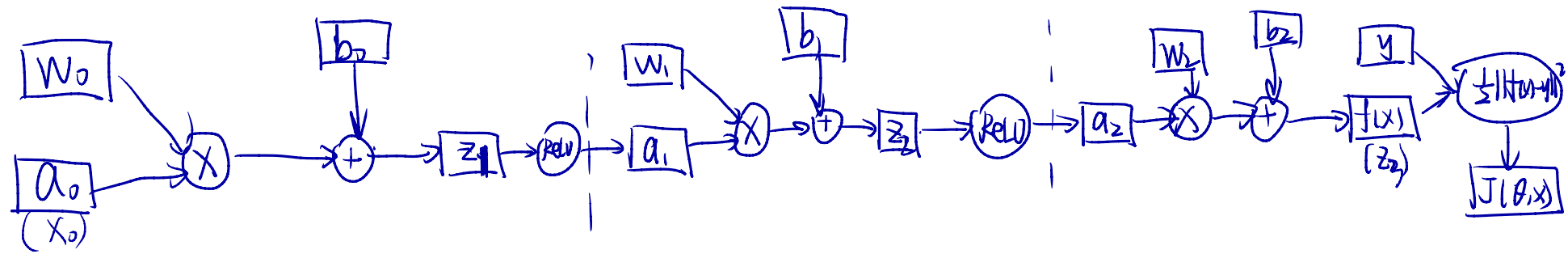


# Problem 1



# Problem 2

$$\frac{\partial J}{\partial y} = -(z_3 - y); \quad \frac{\partial J}{\partial z_3} = z_3 - y; \quad \frac{\partial J}{\partial b_2} = \frac{\partial z_3}{\partial b_2} \frac{\partial J}{\partial z_3} = \frac{\partial J}{\partial z_3}; \quad \frac{\partial J}{\partial W_2} = \frac{\partial J}{\partial z_3} \frac{\partial z_3}{\partial W_2} = a_2^T \frac{\partial J}{\partial z_3};$$

$$\frac{\partial J}{\partial a_2} = \frac{\partial J}{\partial z_3} W_2^T;$$

$$\frac{\partial J}{\partial z_2} = \frac{\partial J}{\partial a_2} \frac{\partial a_2}{\partial z_2} = \frac{\partial J}{\partial a_2} g'(z_2); \quad \frac{\partial J}{\partial b_1} = \frac{\partial J}{\partial z_2} \frac{\partial z_2}{\partial b_1} = \frac{\partial J}{\partial z_2}; \quad \frac{\partial J}{\partial W_1} = \frac{\partial J}{\partial z_2} \frac{\partial z_2}{\partial W_1} = a_1^T \frac{\partial J}{\partial z_2};$$

$$\frac{\partial J}{\partial a_1} = \frac{\partial J}{\partial z_2} \frac{\partial z_2}{\partial a_1} = \frac{\partial J}{\partial z_2} W_1^T$$

$$\frac{\partial J}{\partial z_1} = \frac{\partial J}{\partial a_1} \frac{\partial a_1}{\partial z_1} = \frac{\partial J}{\partial a_1} g'(z_1); \quad \frac{\partial J}{\partial b_0} = \frac{\partial J}{\partial z_1} \cdot 1 = \frac{\partial J}{\partial z_1}; \quad \frac{\partial J}{\partial W_0} = a_0^T \frac{\partial J}{\partial z_1}; \quad \frac{\partial J}{\partial a_0} = \frac{\partial J}{\partial z_1} W_0^T$$

$\uparrow$   
 $\frac{\partial J}{\partial x}$

# Problem 3

```
def compute_gradient(x, y, W0, W1, W2, b0, b1, b2, a1, a2):
    z3=np.dot(a2,W2)+b2
    pz3 = (z3-y)/x.shape[0]  #(500,1)
    pb2 = np.sum(pz3)
    # pb2 = pz3
    pw2 = a2.T.dot(pz3)
    pa2 = pz3.dot(W2.T)
    z2 = np.dot(a1,W1) + b1
    pz2 = relu_derivative(z2)*pa2
    pb1 = np.sum(pz2)
    # pb1 = pz2
    pw1 = a1.T.dot(pz2)
    pa1 = pz2.dot(W1.T)
    a0 = x
    z1 = np.dot(a0,W0)+b0
    pz1 = relu_derivative(z1)*pa1
    pb0 = np.sum(pz1)
    # pb0=pz1
    pw0 = a0.T.dot(pz1)
    return [pw0,pw1,pw2,pb0,pb1,pb2]
```

diff 0.00022914065774079506

diff 0.000819862403225402

diff 0.0035922271467208056

diff 2.364646833952431

diff 4.209797581946187

diff 2.390621034464857e-11

Please make sure all the difference are sufficiently small to go on

0: loss is 7.210390505635944

1: loss is 5.435168663617605

2: loss is 4.33535680069788

3: loss is 3.4824026061434132

4: loss is 2.7427876593981857

5: loss is 2.0783099914825405

6: loss is 1.4934118626272284

7: loss is 1.0134039416728988

8: loss is 0.6580997010116421

9: loss is 0.4285155725361531

10: loss is 0.299973992373061

11: loss is 0.23630293188659873

12: loss is 0.20703182164056785

13: loss is 0.19331879022607307

14: loss is 0.18586068337217101

15: loss is 0.18081173795624347

16: loss is 0.17671770535823414

17: loss is 0.17306318328438955

...

97: loss is 0.08559276882420687

98: loss is 0.0852449544033859

99: loss is 0.08490338145455646

Test loss is 0.11023422312610816