HW 4 Neural Networks

T1

In [1]:
```python
x0 = 1.0
w0 = 0.3; w1 = -0.2
b0 = 0.1; b1 = -0.3
```

In [2]:
```python
def relu(x):
    return max(0, x)

def relu_backward(x):
    return 1 if x > 0 else 0

def forward(x):
    x1 = relu(x0 * w0 + b0) # 0.4
    y1 = x1 * w1 + b1 # -0.08 - 0.3
    z = relu(y1 + x0) # 0.62
    return z, y1, x1

def backward_w0():
    dx1_dw0 = relu_backward(x0 * w0 + b0) * x0
    dy1_dx1 = w1
    dz_dy1 = relu_backward(y1 + x0) * 1
    dz_dw0 = dz_dy1 * dy1_dx1 * dx1_dw0
    return dz_dw0

def backward_w1():
    dy1_dw1 = x1
    dz_dy1 = relu_backward(y1 + x0) * 1
    dz_dw1 = dz_dy1 * dy1_dw1
    return dz_dw1

def backward_b0():
    dx1_db0 = relu_backward(x0 * w0 + b0) * 1
    dy1_dx1 = w1
    dz_dy1 = relu_backward(y1 + x0) * 1
    dz_db0 = dz_dy1 * dy1_dx1 * dx1_db0
    return dz_db0

def backward_b1():
    dy1_db1 = 1
    dz_dy1 = relu_backward(y1 + x0) * 1
    dz_db1 = dz_dy1 * dy1_db1
    return dz_db1

z, y1, x1 = forward(x0)

# z, dz/w0, dz/w1, dz/b0, dz/b1
z, backward_w0(), backward_w1(), backward_b0(), backward_b1()
```

Out[2]:
```
(0.62, -0.2, 0.4, -0.2, 1)
```

T2

In [3]:
```python
batch_size, num_features = 32, 30
neuron_layers = [1024, 512, 1]
```

In [6]:
```python
# input: (30, 32)
A = (1024, 32)
B = (512, 32)
C = (1, 32)

A, B, C
```

Out[6]:
```
((1024, 32), (512, 32), (1, 32))
```

T3

In [7]:
```python
(30 * 1024) + 1024 + (1024 * 512) + 512 + (512 * 1) + 1
```

Out[7]:
```
557057
```

T4

find derivitive of $P(y=j)$ wrt h first

Case $i=j$

$$\frac{\partial P(y=j)}{\partial h_i} = \frac{\partial \frac{\exp(h_i)}{\Sigma_k \exp(h_k)}}{\partial h_i}$$

$$= \frac{\exp(h_i)}{\Sigma_k \exp(h_k)}\left(1 - \frac{\exp(h_i)}{\Sigma_k \exp(h_k)}\right)$$

case $i \neq j$

$$\frac{\partial P(y=j)}{\partial h_i} = \frac{\partial \frac{\exp(h_j)}{\Sigma_k \exp(h_k)}}{\partial h_i}$$

$$= -\frac{\exp(h_i)}{\Sigma_k \exp(h_k)} \cdot \frac{\exp(h_j)}{\Sigma_k \exp(h_k)}$$

$$\frac{\partial L}{\partial P(y=j)} = \frac{-\Sigma_j y_j \log\left(\frac{\exp(h_j)}{\Sigma_k \exp(h_k)}\right)}{\partial\left(\frac{\exp(h_j)}{\Sigma_k \exp(h_k)}\right)}$$

$$= \frac{-\Sigma_j y_j}{\frac{\exp(h_j)}{\Sigma_k \exp(h_k)}}$$

find derivitive of L wrt h after

case $i=j$

$$\frac{\partial L}{\partial h} = \frac{\partial L}{\partial P(y=j)} \cdot \frac{\partial P(y=j)}{\partial h}$$

$$= -\Sigma_i y_i \left(1 - \frac{\exp(h_i)}{\Sigma_k \exp(h_k)}\right)$$

$$= y_i \frac{\exp(h_i)}{\Sigma_k \exp(h_k)} - y_j$$

$$= p_i - y_i$$

case $i \neq j$

$$\frac{\partial L}{\partial h} = -y_i\left(1 - \frac{\exp(h_i)}{\Sigma_k \exp(h_k)}\right) - \Sigma_{j \neq i} y_j \frac{1}{\frac{\exp(h_j)}{\Sigma_k \exp(h_k)}}\left(-\frac{\exp(h_i)}{\Sigma_k \exp(h_k)} \cdot \frac{\exp(h_j)}{\Sigma_k \exp(h_k)}\right)$$

$$= (-y_i + p_i) + \Sigma_{j \neq i} y_j \frac{\exp(h_i)}{\Sigma_k \exp(h_k)}$$

$$= -y_i + p_i \Sigma_j y_j$$

$$= p_i - y_i$$