ReLU

$$f(x) = x^+ = max(x, 0) \tag{1}$$

Softmax

$$f(x_i) = \frac{exp(x_i)}{\sum_j exp(x_j))} \tag{2}$$

Forward Pass

Let $(x_1, y_1, ..., (x_{n_b}, x_{n_b})$ be the data in a mini-batch $\mathcal{D}^{(t)}$, where $X \in R^{d \times n}$ and $Y \in R^{o \times n}$.

for $i = 0$

$$H^0 = ReLU(W_i X_{batch} + b_i 1_{n_b}^T, 0) \tag{3}$$

for $i = 1, ..., k-1$

$$H^i = ReLU(W_i H^{(i-1)} + b_i 1_{n_b}^T, 0) \tag{4}$$

Then,

$$P_{batch} = Softmax(W_k H^{(k-1)} + b_k 1_{n_b}^T) \tag{5}$$

Backward Pass

$$G_{batch} = -(Y_{batch} - P_{batch}) \tag{6}$$

for l = k, k - 1, ..., 2

$$\frac{\partial L}{\partial W_l} = \frac{1}{n_b} G_{batch} H^{(l-1)^T} \tag{7}$$

$$\frac{\partial L}{\partial b_l} = \frac{1}{n_b} G_{batch} 1_{n_b} \tag{8}$$

$$G_{batch} = W_l^T G_{batch} \tag{9}$$

$$G_{batch} = G_{batch} \odot Ind(X_{batch}^{l-1} > 0) \tag{10}$$

Then,

$$\frac{\partial L}{\partial W_1} = \frac{1}{n_b} G_{batch} X_{batch}^T \tag{11}$$

$$\frac{\partial L}{\partial b_1} = \frac{1}{n_b} G_{batch} 1_{n_b} \tag{12}$$