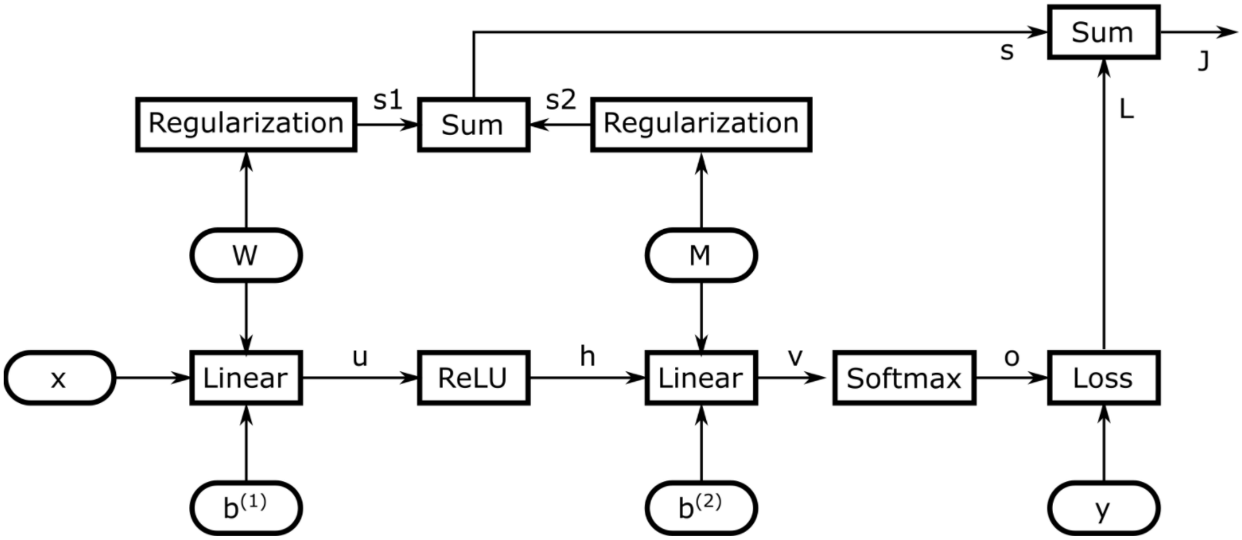
Lab3: Training a Network Using PyTorch's AutoGrad

Author: Nigel Nelson

Date: 3/22/22



$$u = W^T x + b^{(1)}$$

Let $f(x) = x * (x > 0)$

$$h = f(W^T x + b^{(1)})$$

$$v = M^T(f(W^T x + b^{(1)}) + b^{(2)}$$

Let $\sigma(z)_i = \dfrac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$

$$o = \sigma(M^T(f(W^T x + b^{(1)}) + b^{(2)})_i$$

Let loss function $= l(x)$

$$L = l(\sigma(M^T(f(W^T x + b^{(1)}) + b^{(2)})_i, y)$$

$$s1 = \|W\|_F^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \|W_{ij}\|^2$$

where $W \in \mathbb{R}^{m \times n}$

$$s2 = \|M\|_F^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \|M_{ij}\|^2$$

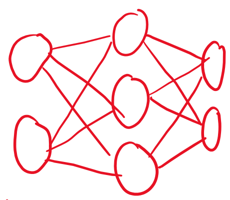
where $M \in \mathbb{R}^{m \times n}$

$$s = \lambda (\|W\|_F^2 + \|M\|_F^2)$$

regularization constant

$$\boxed{J = l(\sigma(M^T(f(W^T x + b^{(1)}) + b^{(2)})_i, y) + \lambda(\|W\|_F^2 + \|M\|_F^2)}$$

Test Cases



$X = 2\times1$ $\quad b_1 = 3\times1$

$W_1 = 3\times2$

$W_2 = 2\times3$ $\quad b_2 = 2\times1$

Case #1

$$x = \begin{bmatrix}0\\1\end{bmatrix} \quad W = \begin{bmatrix}1&1\\1&1\\0&0\end{bmatrix} \quad b_1 = \begin{bmatrix}0\\1\\1\end{bmatrix} \quad y = \begin{bmatrix}2\\2\end{bmatrix}$$

$$M = \begin{bmatrix}1&0&1\\1&0&1\end{bmatrix} \quad b_2 = \begin{bmatrix}1\\0\end{bmatrix}$$

$$u = \begin{bmatrix}1&1\\1&1\\0&0\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix} + \begin{bmatrix}0\\1\\1\end{bmatrix} = \begin{bmatrix}1\\1\\0\end{bmatrix} + \begin{bmatrix}0\\1\\1\end{bmatrix} = \begin{bmatrix}1\\2\\1\end{bmatrix}$$

$$S_1 = \sum \|W\|^2 = 1^2+1^2+1^2+1^2+0^2+0^2 = 4$$

Let $f(x) = x^*(x>0)$

$$h = f(u) = \begin{bmatrix}1\\2\\1\end{bmatrix}$$

$$v = \begin{bmatrix}1&0&1\\1&0&1\end{bmatrix}\begin{bmatrix}1\\2\\1\end{bmatrix} + \begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}2\\2\end{bmatrix} + \begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}3\\2\end{bmatrix}$$

$$S_2 = \sum \|M\|^2 = 1^2+0^2+1^2+1^2+0^2+1^2 = 4$$

$$L = \frac{1}{n}\sum (y-v)^2 = \frac{1}{2}\left((2-3)^2 + (2-2)^2\right) = \frac{1}{2}(1^2+0^2) = \frac{1}{2}(1) = \frac{1}{2}$$

$$S = \lambda(S_1+S_2) = 0.01(4+4) = 0.01(8) = 0.08$$

$$J = L+S = \frac{1}{2} + 0.08 = \boxed{0.58}$$

Case #2

$$x = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \quad W = \begin{bmatrix} 3 & 1 \\ 2 & 4 \\ 0 & 2 \end{bmatrix} \quad b_1 = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} \quad y = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$M = \begin{bmatrix} -2 & 0 & 1 \\ 3 & 2 & 3 \end{bmatrix} \quad b_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

$$u = Wx + b_1 \quad \begin{bmatrix} 3 & 1 \\ 2 & 4 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 9 \\ 16 \\ 6 \end{bmatrix} + \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 11 \\ 14 \\ 7 \end{bmatrix}$$

let $f(x) = x^*(x>0)$

$$h = f(u) = f\left(\begin{bmatrix} 11 \\ 14 \\ 7 \end{bmatrix}\right) = \begin{bmatrix} 11 \\ 14 \\ 7 \end{bmatrix}$$

$$v = Mh + b_2 = \begin{bmatrix} -2 & 0 & 3 \\ 3 & 2 & 3 \end{bmatrix} \begin{bmatrix} 11 \\ 14 \\ 7 \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 82 \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} -2 \\ 84 \end{bmatrix}$$

$$L = \frac{1}{n} \sum (v-y)^2 = \frac{1}{2}\left(\left((1)-(-2)\right)^2 + (2-84)^2\right)$$

$$= \frac{1}{2}\left(3^2 + -82^2\right)$$

$$= \frac{1}{2}(9 + 6{,}724)$$

$$= \frac{1}{2}(6{,}733)$$

$$= 3{,}366.5$$

$$J = L + S = 3{,}366.5 + 0.69 = \boxed{3{,}367.19}$$

$$s_1 = \sum \|W\|^2 = 3^2 + 1^2 + 2^2 + 4^2 + 0^2 + 2^2$$

$$= 9 + 1 + 4 + 16 + 0 + 4$$

$$= 34$$

$$s_2 = \sum \|M\|^2 = -2^2 + 0^2 + 3^2 + 3^2 + 2^2 + 3^2$$

$$= 4 + 9 + 9 + 4 + 9$$

$$= 35$$

$$S = \lambda(s_1 + s_2) = 0.01(34 + 35)$$

$$= 0.01 \cdot 69$$

$$= 0.69$$

Lab3: Training a Network Using PyTorch's AutoGrad
Author: Nigel Nelson
Date: 3/22/22

Case #3

$$x = \begin{bmatrix} -3 \\ 4 \end{bmatrix} \quad W = \begin{bmatrix} -1 & 3 \\ 3 & 2 \\ 4 & -2 \end{bmatrix} \quad b_1 = \begin{bmatrix} -3 \\ 4 \\ 0 \end{bmatrix} \quad y = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

$$M = \begin{bmatrix} 3 & 2 & 1 \\ -1 & -2 & -3 \end{bmatrix} \quad b_2 = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$$

$$u = Wx + b_1 = \begin{bmatrix} -1 & 3 \\ 3 & 2 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} -3 \\ 4 \end{bmatrix} + \begin{bmatrix} -3 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 15 \\ -1 \\ -20 \end{bmatrix} + \begin{bmatrix} -3 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 12 \\ 3 \\ -20 \end{bmatrix}$$

Let $f(x) = x * (x > 0)$

$$h = f(u) = f\left(\begin{bmatrix} 12 \\ 3 \\ -20 \end{bmatrix} \right) = \begin{bmatrix} 12 \\ 3 \\ 0 \end{bmatrix}$$

$$S_1 = \sum ||W||^2 = (-1)^2 + 3^2 + 3^2 + 2^2 + 4^2 + (-2)^2$$
$$= 1 + 9 + 9 + 4 + 16 + 4$$
$$= 43$$

$$V = Mh + b_2 = \begin{bmatrix} 3 & 2 & 1 \\ -1 & -2 & -3 \end{bmatrix} \begin{bmatrix} 12 \\ 3 \\ 0 \end{bmatrix} + \begin{bmatrix} 5 \\ 4 \end{bmatrix} = \begin{bmatrix} 42 \\ -18 \end{bmatrix} + \begin{bmatrix} 5 \\ 4 \end{bmatrix} = \begin{bmatrix} 47 \\ -14 \end{bmatrix}$$

$$S_2 = \sum ||M||^2 = 3^2 + 2^2 + 1^2 + (-1)^2 + (-2)^2 + (-3)^2$$
$$= 9 + 4 + 1 + 1 + 4 + 9$$
$$= 28$$

$$L = \frac{1}{n} \sum (V - y)^2 = \frac{1}{2} \left((47 - 4)^2 + (-14 - 2)^2 \right)$$
$$= \frac{1}{2} \left(43^2 + (-16)^2 \right)$$
$$= \frac{1}{2} (1{,}849 + 256)$$
$$= \frac{1}{2} (2{,}105)$$
$$= 1{,}052.5$$

$$S = \lambda (S_1 + S_2) = 0.01 (43 + 28)$$
$$= 0.01 \cdot 71$$
$$= 0.71$$

$$J = L + S = 1{,}052.5 + 0.71 = \boxed{1{,}053.21}$$