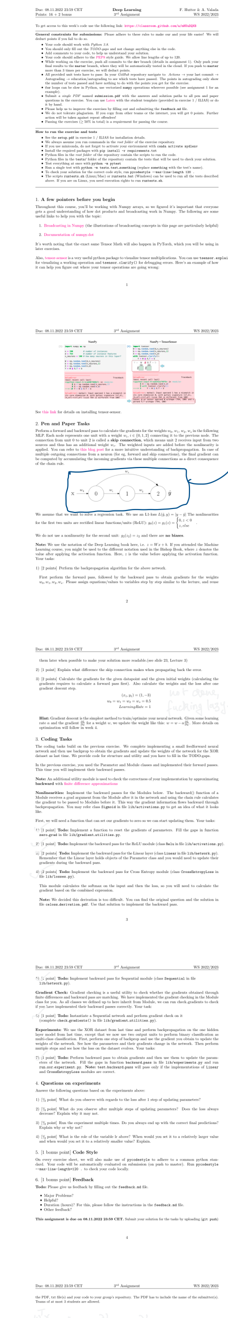
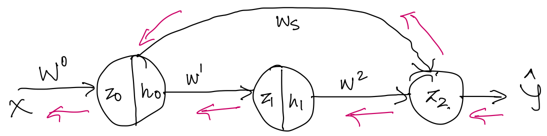
Exercise 3

Saturday, 15. April 2023 15:32

assignm...





Forward pass:

Unit 0

$$z^0 = w^{0T} x + b^0$$
$$h^0 = g^0(z^0)$$

$= 0$

Unit 1

$$z^1 = w^{1T} h^0 + b^1$$
$$h^1 = g^1(z^1)$$

Unit 2

$$z^2 = w^{2T} h^1 + w^{sT} h^0$$
$$\hat{y} = z^2$$

Loss

$$L(y, \hat{y}) = |y - \hat{y}|$$

Backward pass

1) $\frac{\partial L}{\partial z_2} : |y - z_2| = -\frac{(y - z_2)}{|y - z_2|} = L'$

2) $\frac{\partial L}{\partial w_s} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_s} = -\frac{(y-z_2)}{|y-z_2|} h_0 = L' h_0$

$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_2} = -\frac{(y - z_2)}{|y - z_2|} \cdot h_1 = L' h_1$

$\frac{d}{dw}(w^t h)$

$= [w_1 \; w_2 \; w_3] \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}$

$= \boxed{w_1 h_1 + w_2 h_2 + w_3 h_3}$

$= [h_1 \; h_2 \; h_3]$

$M \times N \rightarrow N \times M$

$\frac{\partial L}{\partial h_0} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial h_0} = -\frac{(y - z_2)}{|y - z_2|} w_s = L' w_s$

$\frac{\partial L}{\partial h_1} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial h_1} = -\frac{(y - z_2)}{|y - z_2|} w_2 = L' w_2$

3) $\frac{\partial L}{\partial g_1} = \frac{\partial L}{\partial h_1} \cdot \frac{\partial h_1}{\partial g_1} = L' w_2 \cdot (1)$

$\frac{\partial L}{\partial z_1} = \frac{\partial L}{\partial h_1} \cdot \frac{\partial h_1}{\partial z_1} = L' w_2 \cdot \frac{\partial h_1}{\partial g_1} \cdot \frac{\partial g_1}{\partial z_1} = L' w_2 \cdot 1 \cdot \mathrm{Relu}'(z_1) = L^{z_1}$

4) $z_1 = w_1' h_0$

$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1} = L^{z_1} \cdot h_0$

$\frac{\partial L}{\partial h_0} = \frac{\partial L}{\partial z_1} \cdot \frac{\partial z_1}{\partial h_0} = L^{z_1} \cdot w_1$

Due to skip connection, the gradients of both add up.

$\therefore \frac{\partial L}{\partial h_0} = L^{z_1} w_1 + L' w_s$

5) $h_0 = g_0(z_0)$

$\frac{\partial L}{\partial g_0} = \frac{\partial L}{\partial h_0} \cdot \frac{\partial h_0}{\partial g_0} = \boxed{} \cdot 1$

$\frac{\partial L}{\partial z_0} = \frac{\partial L}{\partial h_0} \cdot \frac{\partial h_0}{\partial g_0} \cdot \frac{\partial g_0}{\partial z_0} = \boxed{} \cdot 1 \cdot \mathrm{Relu}'(z_0)$

6) $z_0 = w^0 x$

$\frac{\partial L}{\partial w^0} = \frac{\partial L}{\partial z_0} \cdot \frac{\partial z_0}{\partial w^0} = \boxed{} x$

$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial z_0} \cdot \frac{\partial z_0}{\partial x} = \boxed{} \cdot w_0$

Final

$\frac{\partial L}{\partial w_0} = (L^{z_1} w_1 + L' w_s)(\mathrm{Relu}'(z_0))(x)$

$\frac{\partial L}{\partial w_1} = L' w_2 \cdot 1 \cdot \mathrm{Relu}'(z_1)$

$\frac{\partial L}{\partial w_2} = \frac{-(y - z_2)}{|y - z_2|} \cdot h_1$

$\frac{\partial L}{\partial w_s} = -\frac{(y - z_2)}{|y - z_2|} h_0$

Backpropagation:

Relu: $\max(0, x)$

Relu' $= \max\left(0, \frac{x}{|x|}\right)$ $\frac{-100}{100}$ $\frac{-100}{100}$

Linear $= w^T x + b$

$\mathrm{Linear}'_w = x^T$

$\mathrm{Linear}'_x = w$

$\mathrm{Linear}'_b = 1$

CE loss with preapplied softmax.

CE loss $= \sum [a \log \mathrm{softmax}(\hat{y}) + (1-a) \log(1 - \mathrm{softmax}(\hat{y}))]$

$\mathrm{softmax}(y) = \sum_{i=1}^{N} \; \exp x^{(i)} =$

$\mathrm{softmax}' = \frac{u}{v} = \frac{vu' - uv'}{v^2} = \frac{\sum \exp(x) \cdot \exp(x_n) - \exp(x_n)^2}{[\sum \exp(x)]^2}$

$j=1$ $\quad \Sigma \mathrm{Exp}(x)^k$

$$(\mathrm{CEloss})' = a\frac{1}{\mathrm{softmax}(\hat{y})} \cdot \mathrm{softmax}(\hat{y})' + (1-a)\frac{1}{1-\mathrm{softmax}(\hat{y})} \cdot \mathrm{softmax}(\hat{y})'$$

$$= \left[\frac{a}{\mathrm{softmax}(\hat{y})} - \frac{(1-a)}{1-\mathrm{softmax}(\hat{y})}\right] \mathrm{softmax}(\hat{y})'$$

$$= \left[\frac{a(1-\mathrm{softmax}) + (a-1)\mathrm{softmax}}{\mathrm{soft}(1-\mathrm{soft})}\right] \mathrm{soft}'(\hat{y})$$

$$= \left[\frac{a - \cancel{a\,\mathrm{soft}} + \cancel{a\,\mathrm{sof}} - \mathrm{soft}}{\mathrm{soft}(1-\mathrm{soft})}\right] \mathrm{soft}'(\hat{y})$$

$$= \Sigma \left[\frac{a - \mathrm{soft}}{\mathrm{soft}(1-\mathrm{soft})}\right] \mathrm{soft}'(\hat{y})$$

$$- \frac{\left[\quad \mathrm{xp}(\ )\right]}{\exp(x_k)\left[\Sigma \exp(x) - \exp(x_k)\right]} \over {\left[\Sigma \exp(x)\right]^2}$$

$$= \frac{\Sigma\left[a_j - \frac{\exp(x^j)}{\Sigma \exp(x)}\right] \cdot \left[\frac{\exp(x_k)\left[\Sigma \exp(x) - \exp(x_k)\right]}{\left[\Sigma \exp(x)\right]^2}\right]}{}$$