Example of Python code optimization

* MM - matrix multiplication

$$A \times B = C, \quad A \in \mathbb{R}^{N \times k}, \; B \in \mathbb{R}^{k \times M}$$
$$C \in \mathbb{R}^{N \times M}$$

$$c_{ij} = \sum_k a_{ik} \cdot b_{kj} \qquad \text{- 3 for loops}$$

1) Get rid of loop over k:

$$c_{ij} = (a_{i,:} * b_{:,j}).\mathrm{sum}()$$

element wise operation

2) Get rid of loop over j:

We gonna use Broadcasting

$$a_{i,:} = (a_{i1} \ldots a_{ik}) \in \mathbb{R}^k \qquad \text{- shape} = k$$

First we are gonna squeeze new dimension (a[:, None]). This will give us $(a_{i1} \ldots a_{ik})^T \in \mathbb{R}^{k \times 1}$

Broadcasting will give us:

$$k \underbrace{\begin{bmatrix} a_{i1} & a_{i1} & \cdots & a_{i1} \\ \vdots & \vdots & & \vdots \\ a_{ik} & a_{ik} & & a_{ik} \end{bmatrix}}_{M} * B = \begin{pmatrix} b_{11} a_{i1} & \cdots & b_{1m} a_{i1} \\ \vdots & & \\ b_{k1} a_{ik} & \cdots & b_{km} a_{ik} \end{pmatrix}$$

elem. wise

Sum over rows equals to c_{i1}

$= c_{i,:}$

Resulting expression :

$$c_{i,:} = (a[:, None] * B).sum(dim=0)$$

Chain Rule

$$y = f(g(x))$$

$$x \xrightarrow{g} h \xrightarrow{f} y$$

$g(x) \quad f(h)$

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial f} \cdot \frac{\partial f}{\partial g} \cdot \frac{\partial g}{\partial x}$$

$$\text{in} \xrightarrow{x} (g) \xrightarrow{g} (f) \xrightarrow{f}$$

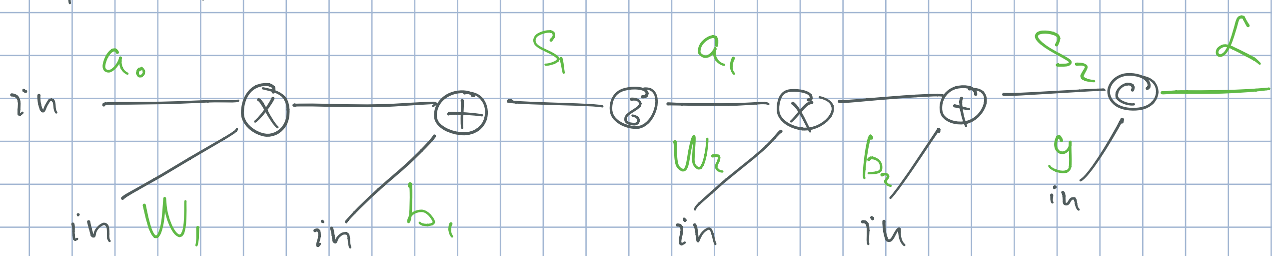
Forward Pass

$s_1 = a_0 \cdot W_1 + b_1 \qquad a_1 = relu(s_1) \quad , \quad a_0 = X \in \mathbb{R}^{1\times n}$

$out = s_2 = a_1 \cdot W_2 + b_2 \quad , \quad \mathcal{L} = (s_2 - y)^2 \in \mathbb{R}$

$W_1 \in \mathbb{R}^{n\times m}, \quad a_1 \in \mathbb{R}^{1\times m}, \quad W_2 \in \mathbb{R}^{m\times 1}$

$b_1 \in \mathbb{R}^{1\times m}$



Backward Pass

$$\frac{\partial \mathcal{L}}{\partial W_1} = \frac{\partial \mathcal{L}}{\partial s_2} \frac{\partial s_2}{\partial a_1} \frac{\partial a_1}{\partial s_1} \frac{\partial s_1}{\partial W_1}$$

$$\frac{\partial y}{\partial x} = \dots$$

input

order of calculation

Shape

$$\frac{\partial \mathcal{L}}{\partial s_2} = 2(s_2 - y) \longrightarrow \text{inp.grad} \qquad 1 \times 1$$

$$\frac{\partial s_2}{\partial a_1} = W_2^T \qquad 1 \times m$$

$$\frac{\partial a_1}{\partial s_1} = \mathbb{1}\{s_1 > 0\} \qquad 1 \times m$$

$$\frac{\partial s_1}{\partial W_1} = a_0 \quad ? \qquad 1 \times n \qquad (m \times n?)$$

$$\frac{\partial s_1}{\partial b_1} = \mathbb{1} \quad ? \qquad 1 \times m$$

$\frac{\partial \mathcal{L}}{\partial W_1}$ should have shape as W_1 i.e. $n \times m$
or $1 \times m$ for b_1

FORWARD

$s_1 = \text{lin}(a_0, W_1, b_1)$

$a_1 = \text{relu}(s_1)$

$s_2 = \text{lin}(a_1, W_2, b_2)$

$\text{out} = \text{mse}(s_2, y)$

BACKWARD

in. out.

mse_grad(s_2, y)

lin_grad(a_1, s_2)

relu_grad(s_1, a_1)

lin_grad(a_0, s_1)

mse_grad(inp, target):

inp.g = 2(inp − target) // inp.g $\triangleq \frac{\partial \mathcal{L}}{\partial \text{inp}}$

! inp.g.shape = inp.shape !

lin_grad(inp, out, W, b): // out.g − already computed

inp.g = out.g @ W^T

$W.g = \frac{\partial \mathcal{L}}{\partial W}$

$b.g = \frac{\partial \mathcal{L}}{\partial b} =$

relu_grad(inp, out)

inp.g = out.g * (inp > 0)

$s_2.g = 2(s_2 - y) \in \mathbb{R}$

$a_1.g = s_2.g \, W_2^T \in \mathbb{R}^{1 \times m}$

$s_1.g = a_1.g * \mathbb{1}\{s_1 > 0\} \in \mathbb{R}^{1 \times m}$

$\underbrace{a_0}_{= x}.g = s_1.g \, W_1^T \in \mathbb{R}^{1 \times n}$

1) $S = a W$

$a \in \mathbb{R}^{1\times n}$

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

a) $\dfrac{\partial s}{\partial a} = W^T$

$s \in \mathbb{R}^{1\times m}$

$s = [s_1 \; s_2 \; \dots s_m] =$

$$s_k = \sum_{i=1}^{n} a_i \cdot W_{ik}$$

$$s_k = s_k(a_1, a_2, \dots, a_n)$$

$$J_s = \begin{bmatrix} \frac{\partial s_1}{\partial a_1} & \frac{\partial s_1}{\partial a_2} & \dots & \frac{\partial s_1}{\partial a_n} \\ \vdots & & & \\ \frac{\partial s_m}{\partial a_1} & \frac{\partial s_m}{\partial a_2} & \dots & \frac{\partial s_m}{\partial a_n} \end{bmatrix} \qquad \frac{\partial S_k}{\partial a_i} = W_{ik}$$

$$= \begin{bmatrix} W_{11} & W_{21} & \dots & W_{n1} \\ W_{12} & W_{22} & \dots & W_{n2} \\ \vdots & & & \\ W_{1m} & W_{2m} & \dots & W_{nm} \end{bmatrix} = W^T$$

b) ?

$\dfrac{\partial s}{\partial W} =$

$$s_k = s_k(W_{1k}, W_{2k}, \dots, W_{nk})$$

$$J_s = \begin{bmatrix} \frac{\partial s_1}{\partial W_{11}} & \frac{\partial s_1}{\partial W_{21}} & \dots & \frac{\partial s_1}{\partial W_{n1}} \\ \frac{\partial s_m}{\partial W_{1m}} & \frac{\partial s_m}{\partial W_{2m}} & \dots & \frac{\partial s_m}{\partial W_{nm}} \end{bmatrix} \qquad \frac{\partial S_k}{\partial W_{ik}} = a_i$$

$$= \begin{bmatrix} a_1 & a_2 & \dots & a_n \\ a_1 & a_2 & \dots & a_n \\ a_1 & \dots & & a_n \end{bmatrix}$$

2) $a = \mathrm{relu}(s) = [\mathrm{relu}(s_1) \; \cdots \; \mathrm{relu}(s_m)]$

$$\frac{\partial a}{\partial s} = \left[\frac{\partial}{\partial s_1} \mathrm{relu}(s_1) \; \cdots \; \frac{\partial}{\partial s_m} \mathrm{relu}(s_m) \right]$$

$$\frac{\partial}{\partial x} \mathrm{relu}(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases}$$

$$\frac{\partial a}{\partial s} = \mathbb{1}\{s > 0\} \quad (\text{elementwise})$$