Section B

Monday, February 27, 2023 11:10 AM

Backprop Ex #1

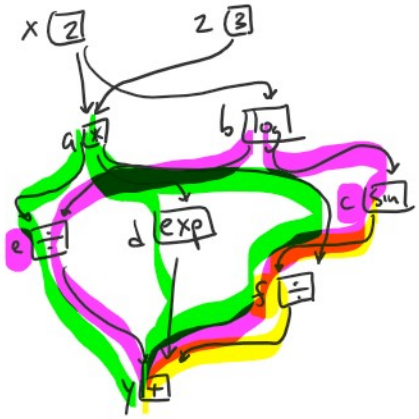
$$y = f(x,z) = \exp(xz) + \frac{xz}{\log(x)} + \frac{\sin(\log(x))}{xz}$$

Forward Computation

Given $x=2$, $z=3$

$a = xz$
$b = \log(x)$
$c = \sin(b)$
$d = \exp(a)$
$e = a/b$
$f = c/a$
$y = d + e + f$

Computation Graph



Backward Computation

$g_y = \frac{dy}{dy} = 1$

$g_f = \frac{dy}{df} = 1$, $g_e = \frac{dy}{de} = 1$, $g_d = \frac{dy}{dd} = 1$

$g_c = \frac{dy}{dc} = \frac{dy}{df}\frac{df}{dc} = (g_f)\left(\frac{1}{a}\right)$

$g_b = \frac{dy}{db} = \frac{dy}{de}\frac{de}{db} + \frac{dy}{dc}\frac{dc}{db}$
$= (g_e)\left(-\frac{a}{b^2}\right) + (g_c)(\cos(b))$

$g_a = \frac{dy}{da} = \frac{dy}{de}\frac{de}{da} + \frac{dy}{dd}\frac{dd}{da} + \frac{dy}{df}\frac{df}{da}$
$= (g_e)\left(\frac{1}{b}\right) + (g_d)(\exp(a)) + (g_f)\left(-\frac{c}{a^2}\right)$

$g_x = \frac{dy}{dx} = (g_a)(z) + (g_b)\left(\frac{1}{x}\right)$

$g_z = \frac{dy}{dz} = (g_a)(x)$

Updates for Backprop

$$g_x = \frac{dy}{dx} = \sum_{k=1}^{K} \frac{dy}{du_k}\frac{du_k}{dx} = \sum_{k=1}^{K} (g_{u_k})\left(\frac{du_k}{dx}\right)$$

Efficient b/c of...

reuse in forward comp.

reuse in backward comp.

Neural Network Training

- Consider a 2-hidden layer NN
- params are $\vec{\theta} = [\alpha^{(1)}, \alpha^{(2)}, \beta]$
- SGD Training:

  Iterate until convergence:

  ① Sample $i \in \{1, \ldots, N\}$:

  ② Compute gradient by backprop:

Background:

$\nabla J(\vec{a},\vec{b}) = \nabla_{\vec{a},\vec{b}} J(\vec{a},\vec{b})$

$\nabla_{\vec{a}} J(\vec{a},\vec{b}) = \begin{bmatrix} dJ/da_1 \\ dJ/da_2 \\ \vdots \\ dJ/da_K \end{bmatrix}$ $|\vec{a}| = K$

(1) Sample $i \in \{1, \dots, N\}$

(2) Compute gradient by backprop:

$$g_{\alpha^{(1)}} = \nabla_{\alpha^{(1)}} J^{(i)}(\vec{\theta})$$
$$g_{\alpha^{(2)}} = \nabla_{\alpha^{(2)}} J^{(i)}(\vec{\theta})$$
$$g_{\beta} = \nabla_{\beta} J^{(i)}(\vec{\theta})$$

$J^{(i)}(\vec{\theta}) = \ell\left(h_\theta(\vec{x}^{(i)}), y^{(i)}\right)$

(3) Update parameters:

$$\alpha^{(1)} \leftarrow \alpha^{(1)} - \gamma g_{\alpha^{(1)}}$$
$$\alpha^{(2)} \leftarrow \alpha^{(2)} - \gamma g_{\alpha^{(2)}}$$
$$\beta \leftarrow \beta - \gamma g_{\beta}$$

Recall: $\vec{\theta} \leftarrow \vec{\theta} - \gamma g_{\vec{\theta}}$

Backprop Ex#2: for NN

Given: (1) Dec. fn. $\hat{y} = h_\theta(\vec{x}) = \sigma\left((\alpha^{(3)})^T \sigma\left((\alpha^{(2)})^T \sigma\left((\alpha^{(1)})^T \vec{x}\right)\right)\right)$

(2) Loss fn. $J = \ell(\hat{y}, y^*) = -\left(y^* \log(\hat{y}) + (1-y^*)\log(1-\hat{y})\right)$

(3) Training ex. $(\vec{x}, y^*)$

leave out intercept terms: β, $\vec{z}^{(2)}$, $\vec{z}^{(1)}$

Forward Comp.

Given $\vec{x}, y^*, \alpha^{(1)}, \alpha^{(2)}, \alpha^{(3)}$

$z^{(0)} = \vec{x}$

for $i = 1, 2, 3$:

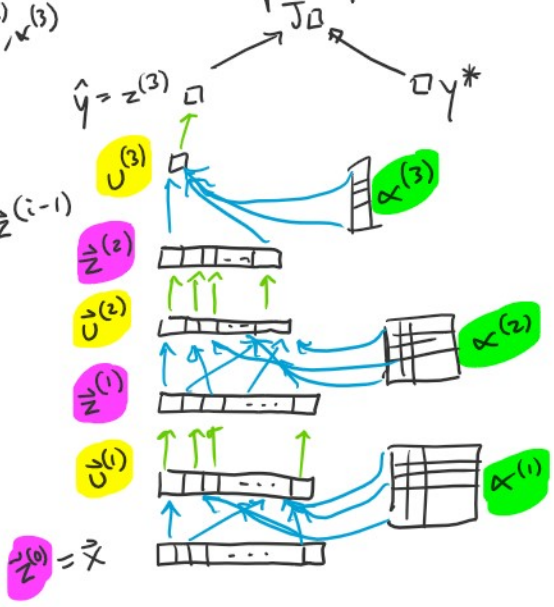
$\vec{v}^{(i)} = (\alpha^{(i)})^T \vec{z}^{(i-1)}$

$\vec{z}^{(i)} = \sigma(\vec{v}^{(i)})$

$\hat{y} = z^{(3)}$

$J = \ell(\hat{y}, y^*)$

Comp. Graph



Backward Comp.

$g_J = [1]$

$g_{\hat{y}} = -\left(\frac{y^*}{\hat{y}} - \frac{(1-y^*)}{(1-\hat{y})}\right)$

for $i = 3, 2, 1$:

$g_{\vec{v}^{(i)}} = \dots$

$g_{\vec{z}^{(i-1)}} = \dots$

$g_{\alpha^{(i)}} = \dots$

$g_{\vec{x}} = g_{\vec{z}^{(0)}}$

HW5

Vector Chain Rule

$y \in \mathbb{R} \quad \vec{v} \in \mathbb{R}^N \quad \vec{x} \in \mathbb{R}^P$

$$\frac{\partial y}{\partial \vec{x}} = \left(\left(\frac{\partial y}{\partial \vec{v}}\right)^T \left(\frac{\partial \vec{v}}{\partial \vec{x}}\right)^T\right)^T$$

$$\left[\frac{\partial y}{\partial x_p}\right] = \sum_{n=1}^{N} \frac{\partial y}{\partial v_n} \frac{\partial v_n}{\partial x_p}$$

$$\frac{\partial y}{\partial \vec{x}} = \left(\left(\frac{\partial y}{\partial \vec{v}} \right)^T \left(\frac{\partial \vec{v}}{\partial \vec{x}} \right)^T \right)^T$$

$P \times 1$; $N \times 1$; $P \times N$; $1 \times P$

$$= \frac{\partial \vec{v}}{\partial \vec{x}} \frac{\partial y}{\partial \vec{v}}$$

$$\left[\frac{\partial y}{\partial \vec{x}} \right]_p = \sum_{n=1}^{N} \frac{\partial y}{\partial v_n} \frac{\partial v_n}{\partial x_p}$$

$$= \sum_{n=1}^{N} \left[\frac{\partial y}{\partial \vec{v}} \right]_n \left[\frac{\partial \vec{v}}{\partial \vec{x}} \right]_{pn}$$