DS-GA 1008: Deep Learning, Spring 2020
Homework Assignment 1

Tejaishwarya Gagadam (tg1779)

February 18, 2020

---

1 Backprop

1.1 Affine Module

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{b}$$

(a) If $C(\mathbf{y})$ is an arbitrary cost function, using chain rule we get,

$$\frac{\partial C}{\partial \mathbf{W}} = \frac{\partial C}{\partial \mathbf{y}} \frac{\partial \mathbf{y}}{\partial \mathbf{W}}$$

$$\frac{\partial C}{\partial \mathbf{b}} = \frac{\partial C}{\partial \mathbf{y}} \frac{\partial \mathbf{y}}{\partial \mathbf{b}}$$

assuming we are given the gradient $\frac{\partial C}{\partial \mathbf{y}}$ and,

$$\frac{\partial \mathbf{y}}{\partial \mathbf{W}} = \mathbf{x}$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{b}} = 1$$

we have,

$$\frac{\partial C}{\partial \mathbf{W}} = \frac{\partial C}{\partial \mathbf{y}} \mathbf{x}$$

$$\frac{\partial C}{\partial \mathbf{b}} = \frac{\partial C}{\partial \mathbf{y}}$$

(b) if have a new cost function $C_2(\mathbf{y}) = 3 * C(\mathbf{y})$, we can calculate the gradients by,

$$\frac{\partial C_2}{\partial \mathbf{W}} = 3 \frac{\partial C}{\partial \mathbf{W}}$$

$$\frac{\partial C_2}{\partial \mathbf{b}} = 3 \frac{\partial C}{\partial \mathbf{b}}$$

which is,

$$\frac{\partial C_2}{\partial \mathbf{W}} = 3\frac{\partial C}{\partial \mathbf{y}}\frac{\partial \mathbf{y}}{\partial \mathbf{W}} = 3\frac{\partial C}{\partial \mathbf{y}}\mathbf{x}$$
$$\frac{\partial C_2}{\partial \mathbf{b}} = 3\frac{\partial C}{\partial \mathbf{y}}\frac{\partial \mathbf{y}}{\partial \mathbf{b}} = 3\frac{\partial C}{\partial \mathbf{y}}$$

We can see that the gradients are scaled by a factor of 3. Hence, we know how our gradients for the new cost function are affected.

1.2 Softmax Module

The derivative of Softmax is,

$$\frac{\partial y_i}{\partial x_j} = \frac{\partial(\frac{e^{\beta x_i}}{\sum_n e^{\beta x_n}})}{\partial x_j}$$

from quotient rule we know that,

$$h(x) = \frac{f(x)}{g(x)}$$
$$h'(x) = \frac{f'(x)g(x) - g'(x)f(x)}{g(x)^2}$$

here,

$$f(x) = e^{\beta x_i}$$
$$g(x) = \sum_n e^{\beta x_n}$$
$$f'(x) = \frac{\partial e^{\beta x_i}}{\partial x_j}$$

if $i = j$

$$f'(x) = \beta e^{\beta x_i}$$

if $i \neq j$

$$f'(x) = 0$$

and,

$$g'(x) = \beta e^{\beta x_j}$$

using the quotient rule, if $i = j$

$$\frac{\partial y_i}{\partial x_j} = \frac{\beta e^{\beta x_i}\sum_n e^{\beta x_n} - \beta e^{\beta x_j} e^{\beta x_i}}{(\sum_n e^{\beta x_n})^2}$$
$$= \frac{\beta e^{\beta x_i}(\sum_n e^{\beta x_n} - e^{\beta x_j})}{(\sum_n e^{\beta x_n})^2}$$

$$= \frac{\beta e^{\beta x_i}}{\sum_n e^{\beta x_n}} \frac{\sum_n e^{\beta x_n} - e^{\beta x_j}}{\sum_n e^{\beta x_n}}$$

$$= \beta y_i (1 - y_j)$$

if $i \neq j$

$$\frac{\partial y_i}{\partial x_j} = \frac{0 - \beta e^{\beta x_j} e^{\beta x_i}}{(\sum_n e^{\beta x_n})^2}$$

$$= -\beta \frac{e^{\beta x_j}}{\sum_n e^{\beta x_n}} \frac{e^{\beta x_i}}{\sum_n e^{\beta x_n}}$$

$$= -\beta y_j y_i$$

So,

$$\frac{\partial y_i}{\partial x_j} = \begin{cases} \beta y_i (1 - y_j) & \text{if } i = j, \\ -\beta y_j y_i & \text{if } i \neq j. \end{cases}$$