Problem Set 1

Author: Alex Enze Liu

Class: TTIC31230 - Fundamentals of Deep Learning

Date: 01/11/2011

1. Problem 1

a)

Let xl denote $\frac{\partial l}{\partial x}$, and according to the question, we know $l = f(y,z) = f\big(g(x),h(x)\big)$

According to calculus, we have.

$$\frac{\partial l}{\partial x} = \frac{\partial f\big(g(x),h(x)\big)}{\partial x} = \frac{\partial f\big(g(x),h(x)\big)}{\partial g(x)} * \frac{\partial g(x)}{\partial x} + \frac{\partial f\big(g(x),h(x)\big)}{\partial h(x)} * \frac{\partial h(x)}{\partial x} = yl * g'(x) + zl * h'(x) \quad \text{(1-1)}$$

Now we calculate $\frac{\partial l}{\partial x}$ based on backpropagation:

Set $x.grad = y.grad = z.grad = 0$

$u.grad = 1$

$$y.grad += u.grad * \frac{\partial f}{\partial y}$$

$$z.grad += u.grad * \frac{\partial f}{\partial z}$$

$$x.grad += y.grad * \frac{\partial g(x)}{\partial x}$$

$$x.grad += z.grad * \frac{\partial h(x)}{\partial x}$$

Calculate the formula above one by one, we have:

$$y.grad = 0 + 1 * \frac{\partial f(y,z)}{\partial y} = \frac{\partial f(y,z)}{\partial y} = yl$$

$$z.grad = 0 + 1 * \frac{\partial f(y,z)}{\partial z} = \frac{\partial f(y,z)}{\partial z} = zl$$

$x.grad = 0 + yl * g'(x)$

$$x.grad = yl * g'(x) + zl * h'(x) = yl * g'(x) + zl * h'(x) \quad \text{(1-2)}$$

We can see that (1-1) is equal to (1-2), which proves back propagation correctly calculates xl.

b)

Let xl denote $\frac{\partial l}{\partial x}$, Let xl' denote $\frac{\partial^2 l}{\partial x^2}$

According to calculus,

$$xl' = \frac{\partial\left(\frac{\partial l}{\partial x}\right)}{\partial x} = \left(\frac{\partial f\big(g(x),h(x)\big)}{\partial g(x)} * g'(x) + \frac{\partial f\big(g(x),h(x)\big)}{\partial h(x)} * h'(x)\right)' = yl' * g'(x)^2 + yl * g''(x) + zl' * h'(x)^2 + zl * h''(x) \tag{1-3}$$

However, according to naïve backpropagation, we have the result

$$yl' * g''(x) + zl' * h''(x) \tag{1-4}$$

We can see that equation (1-3) and (1-4) have different results, which proves $xl' \neq yl' * g''(x) + zl' * h''(x)$

2. Problem 2

Let $\boldsymbol{z}$ denote the input to the softmax function, $\boldsymbol{h}$ denote the output of the softmax function, l denote the loss, we have,

$$\boldsymbol{h} = softmax\ (\boldsymbol{z}) \tag{2-1}$$

We can expand (2-1) as,

$$h_j = \frac{e^{z_j}}{\sum e^{z_i}} \tag{2-2}$$

Based on (2-2), we can have the derivative,

$$\frac{\partial h_i}{\partial z_j} = h_j(1 - h_i) \qquad : i = j \tag{2-3}$$

$$\frac{\partial h_i}{\partial z_j} = -h_i h_j \qquad : i \neq j \tag{2-4}$$

We can combine (2-3) and (2-4) using Kronecker delta,

$$\frac{\partial h_i}{\partial z_j} = h_j\left(\delta_{i,j} - h_i\right) \tag{2-5}$$

Where

$$\delta_{i,j} = \begin{cases} 1\ if\ i = j \\ 0\ if\ i \neq j \end{cases} \tag{2-6}$$

Based on (2-5), we can have,

$$\frac{\partial l}{\partial z_j} = \sum_i \frac{\partial l}{\partial h_i}\frac{\partial hi}{\partial z_j} = \sum_i \frac{\partial l}{\partial h_i} h_j\left(\delta_{i,j} - h_i\right) = \sum_i h_j \left(\frac{\partial l}{\partial h_i}\delta_{i,j} - \frac{\partial l}{\partial h_i} h_i\right) = h_j \sum_i \left(\frac{\partial l}{\partial h_i}\delta_{i,j} - \frac{\partial l}{\partial h_i} h_i\right) \tag{2-7}$$

Based on (2-6) and (2-7), we can have,

$$\frac{\partial l}{\partial z_j} = h_j\left(\frac{\partial l}{\partial h_j} - \sum_i \left(\frac{\partial l}{\partial h_i} h_i\right)\right) \tag{2-8}$$

Based on (2-8), we can have

$$\frac{\partial l}{\partial \boldsymbol{z}} = \boldsymbol{h}\left(\frac{\partial l}{\partial \boldsymbol{h}} - \frac{\partial l}{\partial \boldsymbol{h}} * \boldsymbol{h}\right) \quad \text{(The * here refers to vector dot product)} \tag{2-9}$$

Now let's have a look at the code, we can see that the code and the formula are consistent.

$\frac{\partial l}{\partial \boldsymbol{h}} * \boldsymbol{h}$ refers to the code "np.dot(self.grad, self.value)"

$\boldsymbol{h}\left(\frac{\partial l}{\partial \boldsymbol{h}} - \frac{\partial l}{\partial \boldsymbol{h}} * \boldsymbol{h}\right)$ refers to the code "self.value * (self.grad - gvdot)"

Now we have proved the correctness of the code.