Deep Learning hw wr 2

2018320161 컴퓨터학과 송대선

November 18, 2020

1

$$f(x) = log(\sum_{i=1}^{n} exp(x_i))$$

$$f(x) = log(exp(x_1) + exp(x_2) + \ldots + exp(x_n))$$

$$\bigtriangledown f(x) = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n})$$

$$\bigtriangledown f(x) = (\frac{\partial \sum_{i=1}^{n} exp(x_i)}{\partial x_1}, \frac{\partial \sum_{i=1}^{n} exp(x_i)}{\partial x_2}, \ldots, \frac{\partial \sum_{i=1}^{n} exp(x_i)}{\partial x_n})$$

$$\bigtriangledown f(x) = (\frac{exp(x_1) \times 1}{\sum_{i=1}^{n} exp(x_i)}, \frac{exp(x_2) \times 1}{\sum_{i=1}^{n} exp(x_i)}, \ldots, \frac{exp(x_n) \times 1}{\sum_{i=1}^{n} exp(x_i)})$$

$$\bigtriangledown f(x) = (\frac{exp(x_1)}{\sum_{i=1}^{n} exp(x_i)}, \frac{exp(x_2)}{\sum_{i=1}^{n} exp(x_i)}, \ldots, \frac{exp(x_n)}{\sum_{i=1}^{n} exp(x_i)})$$

$$\therefore \bigtriangledown f(x) = softmax(x)$$

2

$$L(x) = -e_i^T log[softmax(x)]$$

$$L(x) = -e_i^T (log(\frac{(exp(x_1)}{\sum_{k=1}^{n} exp(x_k)}), log(\frac{exp(x_2)}{\sum_{k=1}^{n} exp(x_k)}), \ldots, log(\frac{exp(x_n)}{\sum_{k=1}^{n} exp(x_k)})$$

$$L(x) = -log(\frac{exp(x_i)}{\sum_{k=1}^{n} exp(x_k)}$$

$$L(x) = -log(exp(x_i)) + log(\sum_{k=1}^{n} exp(x_k))$$

$$L(x) = -x_i + log(\sum_{k=1}^{n} exp(x_k))$$

$$\frac{\partial L}{x_i} = -1 + \frac{exp(x_i)}{\sum_{k=1}^{n} exp(x_k)}$$

where $i \neq j$

$$\frac{\partial L}{x_j} = \frac{exp(x_j)}{\sum_{k=1}^{n} exp(x_k)}$$

$$\therefore \bigtriangledown L(x) = \begin{bmatrix} \frac{exp(x_1)}{\sum_{k=1}^{n} exp(x_k)} \\ \vdots \\ -1 + \frac{exp(x_i)}{\sum_{k=1}^{n} exp(x_k)} \\ \vdots \\ \frac{exp(x_n)}{\sum_{k=1}^{n} exp(x_k)} \end{bmatrix}$$

3

$$f(W) = x^T W x$$

$$f(W) = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}$$

$$f(W) = \sum_{j=1}^{n} \sum_{i=1}^{n} x_i w_{ij} x_j$$

$$f_i(W) = \begin{bmatrix} w_{i1} & w_{i2} & \cdots & w_{in} \end{bmatrix}$$

$$\frac{\partial f_i}{\partial w_{ii}} = x_i^2$$

where $i \neq j$

$$\frac{\partial f_i}{\partial w_{ij}} = \frac{\partial f_j}{\partial w_{ij}} = x_i x_j$$

$$\frac{df}{dW} = \begin{bmatrix} \frac{\partial f_1}{\partial w_{11}} & \frac{\partial f_1}{\partial w_{12}} & \cdots & \frac{\partial f_1}{\partial w_{1n}} \\ \frac{\partial f_2}{\partial w_{21}} & \frac{\partial f_2}{\partial w_{22}} & \cdots & \frac{\partial f_2}{\partial w_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial w_{n1}} & \frac{\partial f_n}{\partial w_{n2}} & \cdots & \frac{\partial f_n}{\partial w_{nn}} \end{bmatrix}$$

$$\therefore \frac{df}{dW} = \begin{bmatrix} x_1^2 & x_1x_2 & \cdots & x_1x_n \\ x_1x_2 & x_2^2 & \cdots & x_2x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1x_n & x_2x_n & \cdots & x_n^2 \end{bmatrix}$$

4

Let‘s call node whose input gradient is (d) ”q”

4.a)

$$a = \frac{\partial f}{\partial x}$$

$$a = \frac{\partial q}{\partial x}\frac{\partial f}{\partial q}\frac{\partial f}{\partial f}$$

$$q = xy, \frac{\partial q}{\partial x} = y = 5$$

$$f = qz, \frac{\partial f}{\partial q} = z = 3$$

$$\frac{\partial f}{\partial f} = 1 \; as \; given$$

$$a = \frac{\partial q}{\partial x}\frac{\partial f}{\partial q}\frac{\partial f}{\partial f}$$

$$a = 5 \times 3 \times 1$$

$$\therefore a = 15$$

4.b)

$$b = \frac{\partial f}{\partial y}$$

$$b = \frac{\partial q}{\partial y}\frac{\partial f}{\partial q}\frac{\partial f}{\partial f}$$

$$q = xy, \frac{\partial q}{\partial y} = x = -2$$

$$f = qz, \frac{\partial f}{\partial q} = z = 3$$

$$\frac{\partial f}{\partial f} = 1 \; as \; given$$

$$b = \frac{\partial q}{\partial y}\frac{\partial f}{\partial q}\frac{\partial f}{\partial f}$$

$$b = -2 \times 3 \times 1$$

$$\therefore b = -6$$

4.c)

$$c = \frac{\partial f}{\partial z}$$

$$c = \frac{\partial f}{\partial z}\frac{\partial f}{\partial f}$$

$$f = qz, \frac{\partial f}{\partial z} = q = -10$$

$$\frac{\partial f}{\partial f} = 1 \; as \; given$$

$$c = \frac{\partial f}{\partial z}\frac{\partial f}{\partial f}$$

$$c = -10 \times 1$$

$$\therefore c = -10$$

4.d)

$$d = \frac{\partial f}{\partial q}$$

$$d = \frac{\partial f}{\partial q}\frac{\partial f}{\partial f}$$

$$f = qz, \frac{\partial f}{\partial q} = z = 3$$

$$\frac{\partial f}{\partial f} = 1 \; as \; given$$

$$d = \frac{\partial f}{\partial q}\frac{\partial f}{\partial f}$$

$$d = 3 \times 1$$

$$\therefore d = 3$$

5

$$-\alpha \bigtriangledown L = -\alpha \begin{bmatrix} \frac{\partial L}{\partial x} \\ \frac{\partial L}{\partial y} \end{bmatrix} = -\alpha \begin{bmatrix} 2 + 3y \\ 3x \end{bmatrix}$$

6

(number of filters) = 8 as given
output map width = (input map width - filter width)/stride + 1
$= (32 - 3)/1 + 1 = 30$
output map height = (input map height - filter height)/stride + 1
$= (32 - 3)/1 + 1 = 30$

$\therefore$ (number of filters, output map width, output map height) = (8, 30, 30)