DS-GA 1008: Deep Learning, Spring 2019
Homework Assignment 1

Bingqian Deng

February 15, 2019

1 Backprop

Backpropagation or “backward propagation through errors” is a method which calculates the gradient of the loss function of a neural network with respect to its weights.

1.1 Warm-up

Give an expression for $\frac{\partial L}{\partial W}$ and $\frac{\partial L}{\partial b}$ in terms of $\frac{\partial L}{\partial y}$ and x using the chain rule.

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \quad W = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} W_{11}x_1 + W_{12}x_2 + b_1 \\ W_{21}x_1 + W_{22}x_2 + b_2 \end{bmatrix}$$

$$\frac{\partial L}{\partial y} = \left[\frac{\partial L}{\partial y_1}, \frac{\partial L}{\partial y_2}\right]$$

$$\frac{\partial y}{\partial b} = \begin{bmatrix} \frac{\partial y_1}{\partial b_1} & \frac{\partial y_1}{\partial b_2} \\ \frac{\partial y_2}{\partial b_1} & \frac{\partial y_2}{\partial b_2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_{2\times 2}$$

$$\frac{\partial L}{\partial b} = \begin{bmatrix} \frac{\partial L}{\partial b_1} & \frac{\partial L}{\partial b_2} \end{bmatrix} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial b} = \frac{\partial L}{\partial y} \cdot I = \frac{\partial L}{\partial y}$$

$$\frac{\partial L}{\partial W_{11}} = \frac{\partial L}{\partial y_1} \cdot \frac{\partial y_1}{\partial W_{11}} + \frac{\partial L}{\partial y_2} \cdot \frac{\partial y_2}{\partial W_{11}} = \frac{\partial L}{\partial y_1} \cdot x_1 + \frac{\partial L}{\partial y_2} \cdot 0$$

$$\frac{\partial L}{\partial W_{12}} = \frac{\partial L}{\partial y_1} \cdot \frac{\partial y_1}{\partial W_{12}} + \frac{\partial L}{\partial y_2} \cdot \frac{\partial y_2}{\partial W_{12}} = \frac{\partial L}{\partial y_1} \cdot x_2 + \frac{\partial L}{\partial y_2} \cdot 0$$

$$\frac{\partial L}{\partial W_{21}} = \frac{\partial L}{\partial y_1} \cdot \frac{\partial y_1}{\partial W_{21}} + \frac{\partial L}{\partial y_2} \cdot \frac{\partial y_2}{\partial W_{21}} = \frac{\partial L}{\partial y_1} \cdot 0 + \frac{\partial L}{\partial y_2} \cdot x_1$$

$$\frac{\partial L}{\partial W_{22}} = \frac{\partial L}{\partial y_1} \cdot \frac{\partial y_1}{\partial W_{22}} + \frac{\partial L}{\partial y_2} \cdot \frac{\partial y_2}{\partial W_{22}} = \frac{\partial L}{\partial y_1} \cdot 0 + \frac{\partial L}{\partial y_2} \cdot x_2$$

$$\frac{\partial y}{\partial W} = \left[\begin{array}{cccc} \frac{\partial y_1}{\partial W_{11}} & \frac{\partial y_1}{\partial W_{12}} & \frac{\partial y_1}{\partial W_{21}} & \frac{\partial y_1}{\partial W_{22}} \\ \frac{\partial y_2}{\partial W_{11}} & \frac{\partial y_2}{\partial W_{12}} & \frac{\partial y_2}{\partial W_{21}} & \frac{\partial y_2}{\partial W_{22}} \end{array} \right] = \left[\begin{array}{cccc} x_1 & x_2 & 0 & 0 \\ 0 & 0 & x_1 & x_2 \end{array} \right]$$

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial W} = \frac{\partial L}{\partial y} \left[\begin{array}{cccc} x_1 & x_2 & 0 & 0 \\ 0 & 0 & x_1 & x_2 \end{array} \right]$$

1.2 Softmax

When $i = j$:

$$\frac{\partial y_i}{\partial x_i} = \frac{\partial y_i}{\partial x_j} = \frac{\left(\sum_k \exp\left(\beta x_k\right)\right) (\beta \cdot \exp\left(\beta x_j\right) - \exp\left(\beta x_j\right) \cdot \beta \cdot \exp\left(\beta x_j\right)}{\left(\sum_k \exp\left(\beta x_k\right)\right)^2}$$

$$= \frac{\beta \cdot \exp\left(\beta \cdot x_j\right) \left(\sum_k \exp\left(\beta x_k\right) - \exp\left(\beta x_j\right)\right)}{\left(\sum_k \exp\left(\beta x_k\right)\right)^2}$$

When $i \neq j$:

$$\frac{\partial y_j}{\partial x_i} = \frac{-\beta \cdot \exp\left(\beta x_j\right) \cdot \exp\left(\beta x_i\right)}{\left(\sum_k \exp\left(\beta x_k\right)\right)^2}$$

$$= \frac{-\beta \exp\left(\beta \left(x_j + x_i\right)\right)}{\left(\sum_k \exp\left(\beta x_k\right)\right)^2}$$