Sapienza Università di Roma
Data Science

# Assignment 2

Students name: Giuliano Martinelli 1915652, Gabriele Giannotta 1909375, Mario Dhimitri 1910181

Course: *Advanced Machine Learning* – Professor: *Fabio Galasso*
Due date: *November 19th, 2020*

## Question 2 - Backpropagation

### 2. a

Verify that the loss function defined in Eq. (1) has gradient w.r.t. $z^{(3)}$ as Eq. (2):

$$J\left(\theta, \{x_i, y_i\}_{i=1}^N\right) = \frac{1}{N}\sum_{i=1}^N - \log\left[\frac{\exp\left(z_i^{(3)}\right)_{y_i}}{\sum_{j=1}^K \exp\left(z_i^{(3)}\right)_j}\right] \tag{1}$$

$$\frac{\partial J}{\partial z_i^{(3)}}\left(\theta, \{x_i, y_i\}_{i=1}^N\right) = \frac{1}{N}\left(\psi\left(z_i^{(3)}\right) - \delta_{iy_i}\right) \tag{2}$$

Where $\delta$ is the Kronecker delta:

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}$$

It is possible to verify the initial assumption by calculating the gradient:

1. $f(x) = \frac{1}{N}log(x) \rightarrow \frac{\partial f(x)}{\partial x} = -\frac{1}{Nx}$

   $J\left(\theta, \{x_i, y_i\}_{i=1}^N\right) = f(x) = -\frac{1}{N}log(\psi(z_i^{(3)})_{y_i}), \qquad x = \psi(z_i^{(3)})_{y_i}$

   $\frac{\partial J}{\partial \psi(z_i^{(3)})_{y_i}} = -\frac{1}{N\psi(z_i^{(3)})_{y_i}} = \frac{\partial J}{\partial a_i^{(3)}}$

2. $f(x) = \psi(x) \rightarrow \frac{\partial f(x)}{\partial x} = \psi(x)(1 - \psi(x))$

   $a_i^{(3)} = f(x) = \psi(z_i^{(3)}), \qquad x = z_{y_i}^{(3)}$

   $\frac{\partial a_i^{(3)}}{\partial z_j} = \psi(z_i^{(3)})_{y_i}(1 - \psi(z_i^{(3)})_{y_i})$

3. $\frac{\partial J}{\partial z_i^{(3)}} = \frac{\partial J}{\partial a_i^{(3)}}\frac{\partial a_i^{(3)}}{\partial z_i^{(3)}} = 1(1 - \psi(z_i^{(3)})_{y_i}) = \frac{1}{N}(\psi(z_i^{(3)})_{y_i} - 1)$

   This because $\frac{\partial J}{\partial a_i^{(3)}}$ is the upstream gradient.

## 2. b

To verify that the partial derivative of the loss w.r.t. $W^{(2)}$ is:

$$\frac{\partial J}{\partial W^{(2)}}\left(\theta, \{x_i, y_i\}_{i=1}^N\right) = \sum_{i=1}^{N} \frac{\partial J}{\partial z_i^{(3)}} \cdot \frac{\partial z_i^{(3)}}{\partial W^{(2)}}$$
$$= \sum_{i=1}^{N} \frac{1}{N}\left(\psi\left(z_i^{(3)}\right) - \delta_{iy_i}\right) a_i^{(2)^T}$$

We can use the property as follows, that is:

$f(x) = aW, \qquad \frac{\partial f(x)}{\partial a} = W, \qquad \frac{\partial f(x)}{\partial W} = a$

Using upstream and local gradient:

$\frac{\partial J}{\partial W_2} = \frac{\partial J}{\partial z_i^{(3)}}\frac{\partial z_i^{(3)}}{\partial W_2} \qquad \frac{\partial z_i^{(3)}}{\partial W_2} = a_i^{(2)} \qquad \textit{since} \quad z_i^{(3)} = W_2 a_i^{(2)} + b$

$\frac{\partial J}{\partial W_2} = \frac{1}{N}(\psi z_i^{(3)} - 1)a_i^{(2)}$

To verify that the regularized loss in Eq. (3) has the derivative as Eq. (4):

$$\tilde{J}\left(\theta, \{x_i, y_i\}_{i=1}^N\right) = \frac{1}{N}\sum_{i=1}^{N} -\log\left[\frac{\exp\left(z_i^{(3)}\right)_{y_i}}{\sum_{j=1}^{K}\exp\left(z_i^{(3)}\right)_j}\right] + \lambda\left(\left\|W^{(1)}\right\|_2^2 + \left\|W^{(2)}\right\|_2^2\right) \tag{3}$$

$$\frac{\partial \tilde{J}}{\partial W^{(2)}} = \sum_{i=1}^{N}\frac{1}{N}\left(\psi\left(z_i^{(3)}\right) - \delta_{iy_i}\right) a_i^{(2)^T} + 2\lambda W^{(2)} \tag{4}$$

We can do the following:

$f(x) = \lambda\left(\left\|W^{(1)}\right\|_2^2 + \left\|W^{(2)}\right\|_2^2\right) = \lambda\left(\left\|W^{(1)}\right\|_2^2\right) + \lambda\left(\left\|W^{(2)}\right\|_2^2\right) =$

$\frac{f}{W_2} = 0 + \lambda\frac{\partial(\sum\sum(W_2)^2)}{\partial W_2} = 2\lambda W_2$

$\frac{\partial J}{\partial W_2} = \frac{1}{N}(\psi(z_3) - 1)a_2^T + 2\lambda W_2$

**2. c**

We now derive the expressions for the derivatives of the regularized loss in Eq. (3) w.r.t. W(1), b(1), b(2):

$\frac{\partial J}{\partial z_i^{(3)}} = -\frac{1}{N}(\psi(z_j^{(i)}) - \triangle_i$

$z_i^{(3)} = a_i^{(3)} W^2 + b^2$

$\frac{\partial J}{\partial W_2} = \frac{1}{N}(\psi(z_{ji} - \triangle_i) a_{2i}$

$\frac{\partial J}{\partial a_{2i}} = \frac{1}{N}(\psi(z_{ij} - \triangle_i) W_2$

$\frac{\partial J}{\partial b} = \frac{\partial J}{\partial z_{ji}} = \frac{1}{N}(\psi(z_{jn}) - \triangle_i)$

$\frac{\partial J}{\partial z_{2i}} = \frac{\partial J}{\partial a_{2i}} \frac{\partial J_{2i}}{\partial z_{2i}} \rightarrow a_{2i} \begin{cases} 0, & \text{if } z_{2i} < 0 \\ z_{2i}, & \text{if } z_{2i} \geq 0 \end{cases} \rightarrow \frac{\partial a_{2i}}{\partial z_{2i}} = \begin{cases} 0, & \text{if } z_{2i} < 0 \\ i, & \text{if } z_{2i} \geq 0 \end{cases}$

$\frac{\partial J}{\partial z_{2i}} = \frac{1}{N}(\psi(z_{2i}) - \triangle_i)\delta_i$

$\frac{\partial J}{\partial b_i} = \frac{\partial J}{\partial z_{2i}}$

$\frac{\partial J}{\partial W_1} = \frac{\partial J}{\partial z_{2i}} a_i^{(i)} \quad \| \frac{\partial J}{\partial a_i} = \frac{\partial J}{\partial z_{2i}}$