test

April 9, 2020

Verify that the loss function defined in Eq. (5) has the gradient w.r.t. $z^{(3)}$ as below:

$$\frac{\partial J}{\partial z^{(3)}}(\{x_i, y_i\}_{i=1}^N) = \frac{1}{N}(\psi(z^{(3)}) - \Delta)$$

$$J = \frac{1}{N}\left[\sum_{i=1}^{N} -log\psi(z_{yi})\right] = \frac{1}{N}\left[\sum_{i=1}^{N} -log\frac{e^{z_{y_i}}}{\sum_j e^{z_j}}\right] = \frac{1}{N}\left[\sum_{i=1,j=k_i}^{N} -log\frac{e^{z_{i,j}}}{\sum_j e^{z_j}}\right]$$

$$J = \frac{1}{N}\left[\sum_{i=1}^{N} -log\psi(z_{yi})\right] = \frac{1}{N}\left[\sum_{i=1}^{N} -log\frac{e^{z_{y_i}}}{\sum_j e^{z_j}}\right] = \frac{1}{N}\left[\sum_{i=1,j=k_i}^{N} -log\frac{e^{z_{i,j}}}{\sum_j e^{z_j}}\right]$$

$$\psi(z_{yi}) = \frac{e^{z_{y_i}}}{\sum_j e^{z_j}} = \frac{e^{z_{ik_i}}}{\sum_j e^{z_j}}$$

$$\nabla_{J_{i',j'}} = \frac{\partial J}{\partial z_{i'j'}} = \frac{\partial}{\partial z_{i'j'}}\frac{1}{N}\left[\sum_{i=1}^{N} -log\psi(z_{yi})\right] = -\frac{\partial}{\partial z_{i'j'}}\frac{1}{N}\left[log\psi(z_{yi'})\right] = -\frac{1}{N}\frac{1}{\psi(z_{yi'})}\frac{\partial}{\partial z_{i'j'}}\psi(z_{yi'}) =$$

$$\nabla_{J_{i',j'}} = \frac{\partial J}{\partial z_{i'j'}} = \frac{\partial}{\partial z_{i'j'}}\frac{1}{N}\left[\sum_{i=1}^{N} -log\psi(z_{yi})\right] = -\frac{\partial}{\partial z_{i'j'}}\frac{1}{N}\left[log\psi(z_{yi'})\right] = -\frac{1}{N}\frac{1}{\psi(z_{yi'})}\frac{\partial}{\partial z_{i'j'}}\psi(z_{yi'}) =$$

$$-\frac{1}{N}\frac{1}{\psi(z_{yi'})}\frac{e^{z_{i'j'}}\cdot\delta(j',k_{i'})\sum_j e^{z_{i'j}} - e^{z_{i'k'_i}}e^{z_{i'j'}}}{(\sum_j e^{z_{i'j}})^2} =$$

$$-\frac{1}{N}\frac{\sum_j e^{z_{i'j}}}{e^{z_{i'k'_i}}}\frac{e^{z_{i'j'}}\cdot\delta(j',k_{i'})\sum_j e^{z_{i'j}} - e^{z_{i'k'_i}}e^{z_{i'j'}}}{(\sum_j e^{z_{i'j}})^2} =$$

$$\frac{1}{N}\left[\frac{e^{z_{i'j'}}}{\sum_j e^{z_{i'j}}} - \delta(j',k'_i)\right]$$

1 Question 2c

We can repeatedly apply chain rule as discussed above to obtain the derivatives of the loss with respect to all the parameters of the model $\theta = \left(W^{(1)}, b^{(1)}, W^{(2)}, b^{(2)}\right)$. Derive

the expressions for the derivatives of the regularized loss in Eq. (6) w.r.t. $W^{(1)}, b^{(1)}, b^{(2)}$ **now.**

We start deriving the loss J w.r.t $W^{(1)}$, by applying the chain rule. We'll had then the regularization term. Applying the chain rule we get for $W^{(1)}$ the following:

$$\frac{\partial J}{\partial W^{(1)}} = \frac{\partial J}{\partial z^{(2)}} \cdot \frac{\partial z^{(2)}}{\partial W^{(1)}} \tag{1}$$

$$\tag{2}$$

For $\frac{\partial J}{\partial z^{(2)}}$ we apply iteratively the chain rule obtaining:

$$\begin{aligned} \frac{\partial J}{\partial z^{(2)}} &= \frac{\partial J}{\partial a^{(2)}} \odot \frac{\partial a^{(2)}}{\partial z^{(2)}} && (3) \\ &= \left(\frac{\partial J}{\partial z^{(3)}} \cdot \frac{\partial z^{(3)}}{\partial a^{(2)}} \right) \odot \frac{\partial a^{(2)}}{\partial z^{(2)}} && (4) \\ & && (5) \end{aligned}$$

Where $\odot$ inidicates the element-wise product (Hadamard product) and $\frac{\partial a^{(2)}}{\partial z^{(2)}} \in \mathbb{R}^{10\times 5}$ is the the derivative of the ReLu activation w.r.t. $z^{(2)}$, which is also the heaviside step function:

$$\frac{\partial a_{ij}^{(2)}}{\partial z_{ij}^{(2)}} = \{ cazzatinadentro$$