594 HW3

Casey Sobecks

16 March 2021

1 Problem 1

1.1 Question

In this problem, you will complete the derivation we sketched in class to derive the equations needed to calculate the gradient of the cost function in a feed-forward neural network using backpropagation.
Recall that the output from the j^{th} neuron in layer l is given by:

$$a_i^j = f(z_j^l) \tag{1}$$

where $f(\cdot)$ is our non-linear activation function and

$$z_j^l = \sum_k a_k^{l-1} w_{kj}^j + b_j^l \tag{2}$$

with w_{kj}^l representing the weights and b_j^l and biases of the l^{th} layer of the network. Assume that the cost function $\mathcal{C}$ can be expressed as an average of the squared loss C over each point in the dataset $\mathcal{D}$:

$$\mathcal{C} = \frac{1}{2|\mathcal{D}|} \sum_{\mathbf{x} \in \mathcal{D}} C(\mathbf{x}) = \frac{1}{2} \langle \| a^L(\mathbf{x}) - y(\mathbf{x}) \|^2 \rangle \tag{3}$$

$N \equiv |\mathcal{D}|$ is the number of samples, y is the desired output (target) and $a^L \equiv F(\mathbf{x})$ is the result of the feed-forward neural network which has L layers in total.
Define the notation:

$$\Delta_j^l \equiv \frac{\partial C}{\partial z_j^l} \tag{4}$$

(a) Show that the quantity Δ_j^l can be expressed as:

$$\Delta_j^L = \frac{\partial C}{\partial a_j^l} f'(z_j^L) \tag{5}$$

$$\Delta_j^l = \sum_k \Delta_j^{l+1} [w^{l+1T}]_{kj} f'(z_j^l) \quad l < L \tag{6}$$

(b) Show that the partial derivatives of the cost6 function with respect to the network's weights and biases can be calculated as:

$$\frac{\partial C}{\partial w_{ij}^l} = a_i^{l-1} \Delta_j^l \tag{7}$$

$$\frac{\partial C}{\partial b_j^l} = \Delta_j^l \tag{8}$$

1.2 Part a

For this case, $l = L$. Therefore, this can be solved with a simple chain rule.

$$\Delta_j^L = \frac{\partial C}{\partial z_j^L} = \frac{\partial C}{\partial a_j^L} \frac{\partial a_j^L}{\partial z_j^L}$$

Looking at the second partial derivative, the following can be done.

$$a_j^L = f(z_j^L)$$
$$\frac{\partial a_j^L}{\partial z_j^L} = f'(z_j^L)$$

Therefore, the final solution is

$$\Delta_j^L = \frac{\partial C}{\partial a_j^L} \frac{\partial a_j^L}{\partial z_j^L} = \frac{\partial C}{\partial a_j^L} f'(z_j^L) \tag{9}$$

The next case that is to be considered is for $l < L$.

$$\begin{aligned}
\Delta_j^l = \frac{\partial C}{\partial z_j^l} &= \frac{\partial}{\partial z_j^l} \frac{1}{2} \sum_k (a_k^L - y_k)^2 \\
&= \sum_k (a_k^L - y_k) \frac{\partial}{\partial z_j^l} f(z_k^L) \\
&= \sum_k (a_k^L - y_k) f'(z_k^L) \frac{\partial z_k^L}{\partial z_j^l} \\
&= \Delta_k^l \frac{\partial z_k^L}{\partial z_j^l}
\end{aligned}$$

The next thing to consider is the following

$$z_k^L = \sum_{i=1}^{n_{L-1}} a_i^{L-1} w_{ik}^L + b_k^L = \sum_{i=1}^{n_{L-1}} f(z_k^{L-1}) w_{ik}^L + b_k^L$$

so the following can be written

$$
\begin{aligned}
\frac{\partial z_k^L}{\partial z_j^l} &= \sum_{i=1}^{n_{L-1}} f'(z_i^{L-1}) w_{ik}^L \frac{\partial z_i^{L-1}}{\partial z_j^l} \\
&= \sum_{i_1=1}^{n_{L-1}} \sum_{i_2=1}^{n_{L-2}} f'(z_{i_1}^{L-1}) w_{i_1 k}^L f'(z_{i_2}^{L-2}) w_{i_2 i_1}^{L-1} \frac{\partial z_{i_2}^{L-2}}{\partial z_j^l}
\end{aligned}
$$

Continuing this pattern yields the following,

$$
\begin{aligned}
&= \sum_{i_1=1}^{n_{L-1}} \dots \sum_{i_{L-l}=1}^{n_l} [w_{i_1 k}^L f'(z_{i_1}^{L-1})] \dots [w_{i_{L-l} i_{L-l+1}}^l f'(z_{i_{L-l}}^l)] \frac{\partial z_{i_{L-l}^l}}{\partial z_j^l} \\
&= \sum_{i_1=1}^{n_{L-1}} \dots \sum_{i_{L-l}=1}^{n_l} [w_{i_1 k}^L f'(z_{i_1}^{L-1})] \dots [w_{i_{L-l} i_{L-l+1}}^l f'(z_{i_{L-l}}^l)] \delta_{i_{L-l},j} \\
&= \sum_{i_1=1}^{n_{L-1}} \dots \sum_{j=1}^{n_l} [w_{i_1 k}^L f'(z_{i_1}^{L-1})] \dots [w_{i_j i_{L-l+1}}^l f'(z_{i_j}^l)]
\end{aligned}
$$

Using this general form, the simplest two cases can be checked. The first of these is $l = L - 1$.

$$
\begin{aligned}
\Delta_j^{L-1} &= \sum_{k=1}^{n_L} \Delta_k^L \sum_{i=1}^{n_{L-1}} w_{ik}^L f'(z_i^{L-1}) \delta_{i,j} \\
&= \sum_{k=1}^{n_L} \Delta_k^L [w^{L^T}]_{kj} f'(z_j^l) \\
&= \sum_{k=1}^{n_{l+1}} \Delta_k^{l+1} [w^{(l+1)^T}]_{kj} f'(z_j^l)
\end{aligned}
$$

The next case is $l = L - 2$.

$$
\begin{aligned}
\Delta_j^{L-1} &= \sum_{k=1}^{n_L} \Delta_k^L \sum_{i=1}^{n_{L-1}} w_{ik}^L f'(z_i^{L-1}) \sum_{i'=1}^{n_{L-2}} w_{i'i}^{L-1} f'(z_{i'}^{L-2}) \delta_{i',j} \\
&= \sum_{k=1}^{n_{L-1}} \Delta_k^{l+1} [w^{L-1^T}]_{ij} f'(z_j^{L-2}) \\
&= \sum_{k=1}^{n_{l+1}} \Delta_k^{l+1} [w^{l+1^T}]_{ij} f'(z_j^l)
\end{aligned}
$$

Following this pattern gives the general form.

$$
\Delta_j^l = \sum_{i=1}^{n_{l+1}} \Delta_i^{l+1} [w^{l+1^T}]_{ij} f'(z_j^l)
$$

1.3 Part b

For this problem, the chain rule will be used again.

$$
\frac{\partial C}{\partial w_{ij}^l} = \frac{\partial C}{\partial a_j^l} \frac{\partial a_j^l}{\partial z_i^l} \frac{\partial z_j^l}{\partial w_{ij}^l} = \Delta_j^l \frac{\partial z_j^l}{\partial w_{ij}^l}
$$

Looking at the final partial derivative gives the following.

$$\frac{\partial z_J^l}{\partial w_{ij}^l} = \frac{\partial}{\partial w_{ij}}[a_i^{l-1} w_{ij}^l + b_j^l] = a_i^{l-1}$$

Putting it all together gives,

$$\frac{\partial C}{\partial w_{ij}^l} = \Delta_j^l \frac{\partial z_j^l}{\partial w_{ij}^l} = \Delta_j^l a_i^{l-1}$$

The other portion of this problem is solved the following way.

$$\frac{\partial C}{\partial b_j^l} = \frac{\partial C}{\partial a_j^l} \frac{\partial a_j^l}{\partial z_j^l} \frac{\partial z_j^l}{\partial b_j^l} = \Delta_j^l \frac{\partial z_j^l}{\partial b_j^l}$$

Solving for the last partial,

$$\frac{\partial z_j^l}{\partial b_j^l} = \frac{\partial}{\partial b_j^l}[a_k^{l-1} w_{kj}^l + b_j^l] = 1$$

So the solution is

$$\frac{\partial C}{\partial b_j^l} = \Delta_j^l \frac{\partial z_j^l}{\partial b_j^l} = \Delta_j^l$$

2 Problem 2

2.1 Question

Suppose we replace the our non-linearity $f(z)$ with a linear function $f(z) = z$ at every neuron. Re-derive the backpropagation algorithm for this case.

2.2 Part a

To start this problem, take the derivative of the function $f(z)$.

$$f(z) = z$$
$$f'(z) = 1$$

Therefore, the same equations from problem 1(a) and 1(b) can be used with $f'(z)$ replaced by 1. This gives the following for Part a.

$$\Delta_j^L = \frac{\partial C}{a_j^L}$$
$$\Delta_j^l = \sum_{i=1}^{n_{l+1}} \Delta_i^{l+1} [w^{l+1^T}]_{ij}$$

2.3 Part b

Similarly for part a, $f'(z)$ will be replaced by 1. Since neither of the equations from 1(b) depend on $f'(z)$, the equations for 1(b) are the same equations for this problem.

$$\frac{\partial C}{\partial w_{ij}^{l}} = \Delta_j^l a_i^{l-1}$$
$$\frac{\partial C}{\partial b_j^{l}} = \Delta_j^l$$