Chapter theory 2 solution

Artem Puzanov

2016-08-27

1 Matrix representations of main equations

1.1 part I

Intuition Hadamard product is just an easy way to use vectors instead of diagonal matrices, not much thinking needed here)

Mathematical notation To prove this assertion, we must show that

$$\delta^L = \sum{}'(z^L)\nabla_a C = \nabla_a C \odot \sigma'(z^L)$$

Let's call δ^L for non-Hadamard case δ'^L. Or equivalently, we must prove that

$$\forall i : \delta_i^L = \delta_i'^L$$

, where

$$\delta_i^L = \sigma'(z_i^L)\frac{\partial C}{\partial a_i}$$

Let's write $\delta_i'^L$ in index hell notation. A small note: I shall use I^δ as replacement for $\sum$ notation, as I shall use $\sum$ in it's usual arithmetical sum meaning further.

$$\delta_j'^L = \sum_j I_{ij}^\delta \frac{\partial C}{\partial a_j}$$

By definition, $I_i^\delta = 0$, therefore:

$$\sum_j I_{ij}^\delta \frac{\partial C}{\partial a_j} = I_{ii}^\delta \frac{\partial C}{\partial a_i}$$

Replacing I_{ii}^δ with $\sigma'(z_i^L)$ we get

$$\delta_i'^L = \sigma'(z_i^L)\frac{\partial C}{\partial a_i}$$

Which is what we wanted to prove;

1.2 part II

Intuition Again, nothing hard here. Just prove for any j that result is equivalent to Hadamar project

Mathematical notation Let's call δ^l for non-Hadamard case δ'^l. To prove the statement, we must show that

$$\delta^l = \sum{}'(z^l)(w^{l+1})^T\delta^{l+1} = (w^{l+1})^T\delta^{l+1} \odot \sigma'(z^l)$$

Or equivalently, we must prove that

$$\forall i : \delta_i^l = \delta_i'^l$$

, where

$$\delta_i^l = \sigma'(z_i^l)(\sum_j (w^{l+1})_{ij}^T \delta_j^{l+1})_i$$

A small note: I shall use I^δ as replacement for $\sum$ notation, as I shall use $\sum$ in it's usual arithmetical sum meaning further.

For non-hadamar case

$$\delta_i'^l = \sum_k I_{ik}^\delta (\sum_j (w^{l+1})_{ij}^T \delta_j^{l+1})_k$$

By definition, $I_{i \neq k}^\delta = 0$, therefore:

$$\delta_i'^l = I_{ii}^\delta (\sum_j (w^{l+1})_{ij}^T \delta_j^{l+1})_i$$

Replacing I_{ii}^δ with $\sigma'(z_i^l)$ we get

$$\delta_i'^l = \sigma'(z_i^l)(\sum_j (w^{l+1})_{ij}^T \delta_j^{l+1})_i$$

q.e.d.

1.3 part III

Intuition Not really sure what to show here. Just insert both statements by induction form L to l.

2 Proof of fundamental equations (3) and (4)

2.1 Intuition, equation 3

As stated in the chapter, these equations are just result of derivative-taking rules. Essentialy you have to go *into* C, than *into* a, than into z, where C is loss function of value of σ, σ is activation funtion of z, and z is function of w and b

2.2 Mathematical notation, equation 3

The functional form of cost function is $C(\sigma(z(w,b)))$. We have to prove that

$$\frac{\partial C}{\partial b_i} = \delta_i^l$$

Using chain rule:

$$\frac{\partial C}{\partial b_j} = \frac{\partial C}{\partial \sigma_j} \frac{\partial \sigma_j}{\partial z_j} \frac{\partial z_j}{\partial b_j}$$

$$z_j = \sum_j w_j a_j^{l-1} + b_j$$

$$\frac{\partial z_j}{\partial b_j} = 1$$

than

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

We can notice that by definition of the chain rule (given that σ is a function of z):

$$\frac{\partial C}{\partial \sigma_j} \frac{\partial \sigma_j}{\partial z_j^l} = \frac{\partial C}{\partial z_j^l}$$

Which is precisely the definition of δ_i^l (equation number 29 in the book)

2.3 Intuition, equation 4

We shall use the same trick here, except we have more derivative complexity in z_j^l

2.4 Mathematical notation, equation 4

The functional form of cost function is $C(\sigma(z(w, b)))$. We have to prove that

$$\frac{\partial C}{\partial w_{ij}^l} = \delta_i^l a_j^{l-1}$$

Using chain rule:

$$\frac{\partial C}{\partial w_{ij}} = \frac{\partial C}{\partial \sigma_j} \frac{\partial \sigma_j}{\partial z_j} \frac{\partial z_j}{\partial w_{ij}}$$

By (INSERT NUMBER FROM THE BOOK HERE)

$$z_i = \sum_j w_{ij} a_j^{l-1} + b_i$$

So

$$\frac{\partial z_j}{\partial w_{ij}} = a_j^{l-1}$$

If we put that into our chain rule equation we get:

$$\frac{\partial C}{\partial w_{ij}} = \frac{\partial C}{\partial \sigma_j} \frac{\partial \sigma_j}{\partial z_j} a_j^{l-1}$$

We can notice that by definition of the chain rule (given that σ is a function of z):

$$\frac{\partial C}{\partial \sigma_j} \frac{\partial \sigma_j}{\partial z_j^l} = \frac{\partial C}{\partial z_j^l}$$

Which is precisely the definition of δ_i^l (equation number 29 in the book) Putting that into our previous equations we get:

$$\frac{\partial C}{\partial w_{ij}} = \delta_i^l a_j^{l-1}$$

Which is exactly (BP4)

3 Backpropagation algorithm adaptation

3.1 Intuition, custom activation function

Every neuron is pretty lonely and isolated (too primitive to understand that though), so we just have to replace his personal derivatives, he won't notice.

3.2 Mathematical notation, custom activation function

Let's start with BP1 for arbitrary neuron i:

$$\delta_i^l = \frac{\partial C}{\partial \sigma_j} \sigma'(z_i^l)$$

where σ is an activation function. Replacing that with arbitrarty f gives us

$$\delta_i^l = \frac{\partial C}{\partial f_i} f'(z_i^l)$$

Now for BP2:

$$\delta_i^l = \sigma'(z_i^l)(\sum_j (w^{l+1})_{ij}^T \delta_j^{l+1})_i$$

again, we can replace that with arbitrary f:

$$\delta_i^l = f'(z_i^l)(\sum_j (w^{l+1})_{ij}^T \delta_j^{l+1})_i$$

Those two were easy. Now for interesting cases. Let's try $BP3$ first. By applying chain rule

$$\frac{\partial C}{\partial b_j} = \frac{\partial C}{\partial \sigma_j} \frac{\partial \sigma_j}{\partial b_j}$$

The first multiplier is easy - it's either result of modified BP1 or modified BP2. We'll call it δ_j'. The second one is just $\frac{\partial f_j}{\partial b_j}$. We don't need to use chain rule further, as b_j is a constant. Replacing both multipliers we get:

$$\frac{\partial C}{\partial b_j} = \delta_j' \frac{\partial f_j}{\partial b_j}$$

Now let's go for BP4:

$$\frac{\partial C}{\partial w_{ij}} = \delta_i a_j^{l-1}$$

Let's rewrite it using chain rule

$$\frac{\partial C}{\partial w_{ij}} = \frac{\partial C}{\partial \sigma_j}\frac{\partial \sigma_j}{\partial w_{ij}}$$

Using δ_j' replacement from BP3 solution, and replacing σ with f in the second mulitplier:

$$\frac{\partial C}{\partial w_{ij}} = \delta_j' \frac{\partial f_j}{\partial w_{ij}}$$

We can further expand this eqation by applying chain rule to $\frac{\partial f_j}{\partial w_{ij}}$:

$$\frac{\partial C}{\partial w_{ij}} = \delta_j' \frac{\partial f_j}{\partial z_{ij}} a_j^{l-1}$$

This concludes equations for arbitrary f instead of σ activation function

3.3 Intuition, linear activation function

We can use our previous exercise result to specify BP 1-4 for linear case, by substitution of linear $f(z)$ for arbitrary $f(z)$.

3.4 Mathematical notation, linear activation function

In all our eqautions we simply substitue $f'(z_i^l) = 1$:

BP1:

$$\delta_i^l = \frac{\partial C}{\partial f_i} f'(z_i^l) = \frac{\partial C}{\partial f_i} 1$$

BP2:

$$\delta_i^l = f'(z_i^l)(\sum_j (w^{l+1})_{ij}^T \delta_j^{l+1})_i = 1(\sum_j (w^{l+1})_{ji}^T \delta_j^{l+1})_i$$

BP3:

$$\frac{\partial C}{\partial b_j} = \delta_j' \frac{\partial f_j}{\partial b_j} = \delta_j' 1$$

BP4:

$$\frac{\partial C}{\partial w_{ij}} = \delta_j' 1 a_j^{l-1}$$

This concludes modification of backpropagation equations