Chapter theory 5 solution

Artem Puzanov

2016-09-27

1 Unstable gradient for simple networks

1.1 Futility of different activation value

Intuition The core problem of vanishing/exploding gradient is not value per se, but chain rule. Any consistent increasing/decreasing effect is bound to become a problem if the chain is long enough. Some combinations of weight initialization and activation functions might be able to postpone the problem. The longer the chain, the harder it will be to keep this balance though.

No mathematical notation needed

1.2 Numerical bounds of weighted derivative

Intuition Nothing hard here, just old tricks

Mathematical notation We want to prove that:

$$|w\sigma'(wa+b)| > 1$$

if $|w| > 4$. To do that, let's remember that $\sigma'(z) = \sigma(z)(1-\sigma(z))$. Maximum value for $x(1-x)$ is 0.25. So

$$|w0.25| > 1$$

For that to be true, $|w| > 4$.

1.3 Activation bounds

Intuition Again, nothing interesting

Mathematical notation Using derivative for σ:

$$\sigma'(z) = \frac{e^(-wa-b)}{1+e^{-wa-b}}$$

Setting $e^{-wa-b} = x$, we get:

$$w\frac{x}{1-x} >= 1$$

Solving that for x and assuming $w > 4$ we get:

$$\frac{(|w|-2)-\sqrt{w^2-4|w|}}{2} \leq x \leq \frac{(|w|-2)+\sqrt{w^2-4|w|}}{2}$$

Using $e^{-wa-b} = x$, and making some changes to suit the answer better:

$$\frac{|w|\left(1-\sqrt{1-4/|w|}\right)}{2} - 1 \leq e^{-wa-b} \leq \frac{|w|\left(1+\sqrt{1-4/|w|}\right)}{2} - 1$$

Solving for a:

$$\frac{1}{|w|}\ln\left(\frac{|w|\left(1-\sqrt{1-4/|w|}\right)}{2}-1\right)-\frac{b}{|w|}\leq a\leq\frac{1}{|w|}\ln\left(\frac{|w|\left(1+\sqrt{1-4/|w|}\right)}{2}-1\right)-\frac{b}{|w|}$$

Taking a difference between the larger nad the smaller value (and simplyfing again) gives us the desired interval:

$$\frac{2}{|w|}\ln\left(\frac{|w|(1+\sqrt{1-4/|w|})}{2}-1\right)$$

1.4 Sigmoid neuron to identity neuron

Intuition The key idea is to creatively use coefficient, so that on $0 <= x <= 1$ our function behaves linearly. To do that, we need to make $f'(x)dx = dx$ for our neuron, while keeping $f''(x), f'''(x)$ and so on close to zero.

Mathematical notation The first trick is to center our neuron on $0.5, 0.5$ coordinates. That can be done with setting $w_1 = -2b$, so that $1/(1+e^{w_1x+b}) = 1/2$. (Note: switchting signs in exponent is to save time on all that sign acrobatic) Let's write down taylor expansion to the second order:

$$f(a) \approx f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2$$

For our case that shall be

$$\frac{w_2}{1+e^{w_1a+b}} - \frac{w_1w_2e^{w_1a+b}}{(e^{w_1a+b}+1)^2}(x-a) + \frac{w_1^2w_2e^{w_1a+b}(e^{w_1a+b}-1)}{(e^{w_1a+b}+1)^3}(x-a)^2$$

This unholy thing is actually pretty mallable, once we remember that w_1 and b are small. We need second derivative to go to zero - for that $w_2 < 1/w_1^2$ is enough (check the limits if you want, though it's obvious) Remembering that w_1 and b are small, we get:

$$\frac{w_2}{2} - \frac{w_1 * w_2(x-a)}{4}$$

Now, let's try to make sure that his function is equal to x around $a = 1/2$:

$$\frac{w_2}{2} - \frac{w_1 * w_2(x-1/2)}{4} = \frac{w_2}{2} + \frac{w_1w_2}{8} - w_1w_2x/4$$

By setting $w_2 = -4/w_1$, we can get:

$$-2/w_1 - 1/2 + x = x - 2/w_1 - 1/2$$

. Now, a note: this is definetly not just x, as it's tilted by a constant. So it seems I haven't solved the original :(But! if we assume we can not only use w_2, but also have b_2, and set $b_2 = 2/w_1 + 1/2$, than we get a perfect x. NOTE: return here later.