Patterns

As fun as it is to work out derivatives using deltas like Δx and seeing what happens when we make them smaller and smaller, we can often do it without doing all that work. See if you can see any pattern in the derivatives we've worked out so far:

$$s = t^2 \qquad \longrightarrow \qquad \frac{\delta s}{\delta t} \;=\; 2t$$

$$s = t^2 + 2t \qquad \longrightarrow \qquad \frac{\delta s}{\delta t} \;=\; 2t + 2$$

$$s = t^3 \qquad \longrightarrow \qquad \frac{\delta s}{\delta t} \;=\; 3t^2$$

You can see that the derivative of a function of t is the same function but with each power of t reduced by one. So t^4 becomes t^3, and t^7 would become t^6, and so on. That's really easy! And if you remember that **t** is just t^1, then in the derivative it becomes t^0, which is 1.

Constant numbers on their own like 3 or 4 or 5 simply disappear. Constant variables on their own, which we might call a, b or c, also disappear, because they too have no rate of change. That's why they're called *constants*.

But hang on, t^2 became $2t$, not just t. And t^3 became $3t^2$ not just t^2. Well, there is an extra step where the power is used as a multiplier before it is reduced.

So the 5 in $2t^5$ is used as an additional multiplier before the power is reduced $5 * 2t^4 = 10t^4$. The following summarises this power rule for doing calculus.

$$y = ax^n \longrightarrow \frac{\delta y}{\delta x} = nax^{n-1}$$

Let's try it on some more examples just to practice this new technique.

$$s = t^5 \longrightarrow \frac{\delta s}{\delta t} = 5t^4$$

$$s = 6t^6 + 9t + 4 \longrightarrow \frac{\delta s}{\delta t} = 36t^5 + 9$$

$$s = t^3 + c \longrightarrow \frac{\delta s}{\delta t} = 3t^2$$

So this rule allows us to do quite a lot of calculus and for many purposes, it's all the calculus we need. Yes, it only works for *polynomials*, that is expressions made of variables with powers like $y =$ax^3+bx^2+cx + d, and not with things like $sin(x)$ or $cos(x)$. That's not a major flaw because there are a huge number of uses for doing calculus with this power rule.

However, for neural networks, we do need one extra tool, which we'll look at next.