This document is to collect my thoughts about gradient descent and backpropagation as well as figuring out how to get a generic backpropagation algorithm, without it being fixed on a certain network, on my own.
This example assumes that the layers are numbered from 1 to n, with the first layer being the first hidden layer and the nth layer being the output layer. It also assumes the costfunction $\sum \frac{1}{2}(target - out)^2$ and that the sigmoid function (sig(x)=$\frac{1}{1+e^{-x}}$) is used as an activation function and uses the given example of a 2-4-3-2 network.
The terminology and naming of neurons outputs etc. is mostly from this amazing blogpost and is a prerequisite for understanding the following calculations.

Now if you want to calculate the gradient for a weight of the outputlayer, for example w_{31} connecting neurons c_1 and d_1, this is fairly simple. Weight w_{31} is right in the beginning (or rather at the end of the network, but since we are trying to do backpropagation the end is our beginning), most of the network becomes irrelevant when calculating the (partial) derivative.
Since out intention is to minimize the error we will need to calculate the derivative of the error function E_{total} with respect to w_{31}. When following the chain rule to calculate the derivative we get:

$$\frac{\partial E_{total}}{\partial w_{31}} = \frac{\partial E_{total}}{\partial out_{d1}} * \frac{\partial out_{d1}}{\partial net_{d1}} * \frac{\partial net_{d1}}{\partial w_{31}} \tag{1}$$

Because I had some trouble understanding how the left hand side of the equation equals the right hand side at first and this is essential to understanding anything following this equation, I will explain it in more detail before proceeding.
For that let's split up the Error function into its components. In this particular example we begin with $E_{total} = \sum \frac{1}{2}(target - out)^2$. Having only two output neurons this equals $E_{total} = \frac{1}{2}(target_{d_1} - out_{d_1})^2 + \frac{1}{2} * (target_{d_2} - out_{d_2})^2$. The first part of the sum is the error E_{d_1} for neuron d_1 and the second is the error E_{d_2} for neuron d_2.
We want to derive with respect to w_{31} and not the output of some neuron so lets split those up again. Then we would get $out_{d_1} = sig(net_{d_1})$ and $out_{d_2} = sig(net_{d_2})$.
To have w_{31} in our equation we need to split it up once more and get $net_{d_1} = w_{31} * out_{c_1} + w_{xx} * out_{c_2} + w_{xx} * out_{c_3}$ and $net_{d_2} = w_{xx} * out_{c_1} + w_{xx} * out_{c_2} + w_{xx} * out_{c_3}$. I have omitted stating the correct indices for anything but w_{31} because, as we will see soon, they become irrelevant when deriving

with respect to w_{31}.

Putting it back together we get

$$E_{d_1} = \frac{1}{2}(target_{d_1} - sig(w_{31} * out_{c_1} + w_{xx} * out_{c_2} + w_{xx} * out_{c_3}))^2 \quad (2)$$

When beginning to derive this formula, we already need the chain rule. We derive the outer function and multiply it by the derivative of the inner function resulting in

$$(inner) * \frac{\partial inner}{\partial w_{31}}, \text{ mit } inner = target_{d_1} - sig(w_{31} * out_{c_1} + w_{xx} * out_{c_2} + w_{xx} * out_{c_3}) \quad (3)$$

Now to get the first factor you could also derive the equation

$$E_{d_1} = \tfrac{1}{2}(target_{d_1} - out_{d_1})^2$$

with respect to out_{d_1}, which would be

$$\frac{\partial E_{d_1}}{\partial out_{d_1}} = \frac{\partial \frac{1}{2}(target_{d_1} - out_{d_1})^2}{\partial out_{d_1}} = (target_{d_1} - out_{d_1}) * (-1)$$

The only difference is that out_{d_1} isn't split up into its components and the -1 at the end. The -1 is a leftover from deriving the inner function, which looks like it doesn't exist in equation (3), but it is only hidden in $\frac{\partial inner}{\partial w_{31}}$. Now as you can see, $\frac{\partial E_{d_1}}{\partial out_{d_1}}$ is the first factor in equation one, so we already got that part figured out. To get the second and third factor we now have to continue deriving with $\frac{\partial inner}{\partial w_{31}}$ (see equation 3). Since w_{31} is still nested in a function, in this case sig(x), we need to apply the chain rule a second time, which results in

$$\frac{\partial inner}{\partial w_{31}} = -sig(net_{d_1}) * (1 - sig(net_{d_1})) * \frac{\partial net_{d_1}}{\partial w_{31}}, \quad (4)$$

$$\text{mit } net_{d_1} = w_{31} * out_{c_1} + w_{xx} * out_{c_2} + w_{xx} * out_{c_3}$$

Let's look at the first factor first, which ist pretty much only the derivative or the sigmoid function. Now since the output out_{d_1} is nothing but $sig(net_{d_1})$, the derivative of out_{d_1} is also the derivative of $sig(net_{d_1})$. Therefore we get

$$\frac{\partial out_{d_1}}{\partial w_{31}} = \frac{\partial sig(net_{d_1})}{\partial w_{31}} \quad (5)$$

As seen in equation (4), when we apply the chainrule to $sig(net_{d_1})$ we first derive with respect to net_{d_1} and then multiply by the derivative with respect to w_{31}. This means we now have the equation

$$\frac{\partial sig(net_{d_1})}{\partial w_{31}} = \frac{\partial sig(net_{d_1})}{\partial net_{d_1}} * \frac{\partial net_{d_1}}{\partial w_{31}} \tag{6}$$

From the equality in equation (5) we can deduct that

$$\frac{\partial sig(net_{d_1)}}{\partial w_{31}} = \frac{\partial sig(net_{d_1})}{\partial net_{d_1}} * \frac{\partial net_{d_1}}{\partial w_{31}} = \frac{\partial out_{d_1}}{\partial net_{d_1}} * \frac{\partial net_{d_1}}{\partial w_{31}} = \frac{\partial out_{d_1}}{\partial w_{31}} \tag{7}$$

All those steps put back together we have now proven, that by following the chain rule $\frac{\partial E_{d_1}}{\partial w_{31}} = \frac{\partial E_{total}}{\partial out_{d1}} * \frac{\partial out_{d1}}{\partial net_{d1}} * \frac{\partial net_{d1}}{\partial w_{31}}$ (see equation 1).
While we wanted it for E_{total} and not just E_{d_1} you can see, when looking at E_{d_1}, that it does not contain w_{31} at all, meaning that deriving with respect to w_{31} must result in 0.
In other words, for weight w_{31} applies that

$$\frac{\partial E_{d_1}}{\partial w_{31}} = \frac{\partial E_{total}}{\partial w_{31}} = \frac{\partial E_{total}}{\partial out_{d1}} * \frac{\partial out_{d1}}{\partial net_{d1}} * \frac{\partial net_{d1}}{\partial w_{31}} \tag{8}$$

Formula for deriving the gradient with respect to a weight w_{3x} by taking the example of w_{31}

$$\frac{\partial E_{total}}{w_{31}} = \frac{\partial E_{d_1} + E_{d_2}}{\partial w_{31}} = \frac{\partial E_{d_1}}{\partial w_{31}} + \frac{\partial E_{d_2}}{\partial w_{31}} \tag{9}$$

$$\frac{\partial E_{d_1}}{\partial w_{31}} + \frac{\partial E_{d_2}}{\partial w_{31}} = \frac{\partial E_{d_1}}{\partial w_{31}} + 0 = \frac{\partial E_{d_1}}{\partial w_{31}} \tag{10}$$

$$\frac{\partial E_{d_1}}{\partial w_{31}} = \frac{\partial E_{d_1}}{\partial out_{d_1}} * \frac{\partial out_{d_1}}{\partial w_{31}} = \frac{\partial E_{d_1}}{\partial out_{d_1}} * \frac{\partial out_{d_1}}{\partial net_{d_1}} * \frac{\partial net_{d_1}}{\partial w_{31}} \tag{11}$$

Plugging in the actual values for each factor in equation (11) we get

$$\frac{\partial E_{d_1}}{\partial out_{d_1}} = (out_{d_1} - target_{d_1}) \tag{12}$$

$$\frac{\partial out_{d_1}}{\partial net_{d_1}} = sig'(net_{d_1}) = out_{d_1} * (1 - out_{d_1}) \tag{13}$$

$$\frac{\partial net_{d_1}}{\partial w_{31}} = \frac{\partial w_{31} * out_{c_1} + w_{3x} * out_{c_2} + w_{3x} * out_{c_3}}{\partial w_{31}} = out_{c_1} \tag{14}$$

Resulting in the final equation

$$\frac{\partial E_{total}}{\partial w_{31}} = \frac{\partial E_{d_1}}{\partial w_{31}} = (out_{d_1} - target_{d_1}) * out_{d_1} * (1 - out_{d_1}) * out_{c_1} \tag{15}$$

Formula for deriving the gradient with respect to a weight w_{2x} by taking the example of w_{21}

$$\frac{\partial E_{total}}{\partial w_{2_1}} = \frac{\partial E_{d_1}}{\partial w_{21}} + \frac{\partial E_{d_2}}{w_{21}} \tag{16}$$

Starting with the first summand

$$\frac{\partial E_{d_1}}{\partial w_{21}} = \frac{\partial E_{d_1}}{\partial out_{d_1}} * \frac{\partial out_{d_1}}{\partial net_{d_1}} * \frac{\partial net_{d_1}}{w_{21}} \tag{17}$$

$$\frac{\partial net_{d_1}}{w_{21}} = \frac{\partial w_{31} * out_{c_1} + w_{3x} * out_{c_2} + w_3x * out_{c_1}}{\partial w_{21}} \tag{18}$$

$$\frac{\partial net_{d_1}}{w_{21}} = \frac{\partial w_{31} * out_{c_1}}{\partial w_{21}} + \frac{w_{3x} * out_{c_2}}{\partial w_{21}} + \frac{w_3x * out_{c_3}}{\partial w_{21}} \tag{19}$$

$$\frac{\partial net_{d_1}}{w_{21}} = \frac{\partial w_{31} * out_{c_1}}{\partial w_{21}} + 0 + 0 = \frac{\partial w_{31} * out_{c_1}}{\partial w_{21}} \tag{20}$$

$$\frac{\partial net_{d_1}}{w_{21}} = \frac{\partial net_{d_1}}{\partial out_{c_1}} * \frac{\partial out_{c_1}}{\partial net_{c_1}} * \frac{\partial net_{c_1}}{\partial w_{21}} \tag{21}$$

Putting it together for E_{d_1} we get

$$\frac{\partial E_{d_1}}{\partial w_{21}} = \frac{\partial E_{d_1}}{\partial out_{d_1}} * \frac{\partial out_{d_1}}{\partial net_{d_1}} * \frac{\partial net_{d_1}}{\partial out_{c_1}} * \frac{\partial out_{c_1}}{\partial net_{c_1}} * \frac{\partial net_{c_1}}{\partial w_{21}} \tag{22}$$

Plugging in the values this equals

$$\frac{\partial E_{d_1}}{\partial w_{21}} = (out_{d_1} - target_{d_1}) * out_{d_1} * (1 - out_{d_1}) * w_{31} * out_{c_1} * (1 - out_{c_1}) * out_{b_1} \tag{23}$$

With that we have the first part of our total gradient, now we need the second part $\frac{\partial E_{d_2}}{\partial w_{21}}$.
In step (20) we see that two of the factors result in 0 when derived as they are not influenced by w_{21}. When calculating the second part of the gradient this pattern will repeat itself, but we will one weight will matter for us. It is the connection between c_1 and d_1, lets call it w_{34}. The rest of the process is exactely the same and will result in a similar expression to (22)

$$\frac{\partial E_{d_2}}{\partial w_{21}} = \frac{\partial E_{d_2}}{\partial out_{d_2}} * \frac{\partial out_{d_2}}{\partial net_{d_2}} * \frac{\partial net_{d_2}}{\partial out_{c_1}} * \frac{\partial out_{c_1}}{\partial net_{c_1}} * \frac{\partial net_{c_1}}{\partial w_{21}} \tag{24}$$

Now just like above (23) let's plug in the the values

$$\frac{\partial E_{d_2}}{\partial w_{21}} = (out_{d_2} - target_{d_2}) * out_{d_2} * (1 - out_{d_2}) * w_{34} * out_{c_1} * (1 - out_{c_1}) * out_{b_1} \tag{25}$$

Put these two together and you have the gradient with respect to w_{21}.