We want: $\delta_1 = \frac{\partial C}{\partial z_1}$. (We already have expressions relating δ_1 to the desired gradients $\frac{\partial C}{\partial w_1}$ and $\frac{\partial C}{\partial w_{1b}}$.)

To calculate δ_1, we want to rewrite $\frac{\partial C}{\partial z_1}$ in terms of other partial derivatives that we already have.

Consider that C is a multivariate function over many things; among these, C is a function over both z_{21} and z_{22}.

z_{21} and z_{22}, in turn, can <u>each</u> be written as a function of z_1.

$$C = f(z_{21}, z_{22}) \qquad z_{21} = g_1(z_1)$$
$$z_{22} = g_2(z_1)$$

If you recall your multivariate chain rule, this means we can rewrite

$$C = f(g_1(z_1), g_2(z_1))$$

$$\frac{\partial C}{\partial z_1} = \frac{\partial C}{\partial g_1} \cdot \frac{\partial g_1}{\partial z_1} + \frac{\partial C}{\partial g_2} \cdot \frac{\partial g_2}{\partial z_1}$$

Of course, we actually know the functions g_1 and g_2. (These are just the expressions relating z_1 to z_{21} and z_1 to z_{22}, respectively.)

So, we get: $$\frac{\partial C}{\partial z_1} = \frac{\partial C}{\partial z_{21}} \cdot \frac{\partial z_{21}}{\partial z_1} + \frac{\partial C}{\partial z_{22}} \cdot \frac{\partial z_{22}}{\partial z_1}$$

Notice that the first element in each term here is just the δ value we already know: $\frac{\partial C}{\partial z_{21}} = \delta_{21}$ and $\frac{\partial C}{\partial z_{22}} = \delta_{22}$.

All that is left are $\frac{\partial z_{21}}{\partial z_1}$ and $\frac{\partial z_{22}}{\partial z_1}$, which are super easy since we know that:

$$z_{21} = w_{21b} + w_{21}\,\sigma_1(z_1)$$
$$z_{22} = w_{22b} + w_{22}\,\sigma_1(z_1)$$

So, $\frac{\partial z_{21}}{\partial z_1} = w_{21} \cdot \sigma_1'(z_1)$ and $\frac{\partial z_{22}}{\partial z_1} = w_{22} \cdot \sigma_1'(z_1)$