Task 1.1

- We were a bit confused about whether we were supposed to prove BP1 or BP4 (or both) from Nielsen's book, so we did both:

BP4: assume squared error (it's possible to prove with a general cost function, but we thought this to be sufficient):

$$\frac{\partial C}{\partial w_{ji}} = \frac{\frac{1}{2}\sum_k (a_k - t_k)}{\partial w_{ji}} = \sum_k (a_k - t_k)\frac{\partial a_k}{\partial w_{ji}} \quad \text{(chain rule)}$$

$$= \sum_k (a_k - t_k)\frac{\partial f(z_k)}{\partial w_{ji}} \Rightarrow z_k = a_j w_{kj} = \frac{\partial f(z_k)}{\partial w_{ji}} = \cancel{\quad} f'(z_j) w_{kj} \frac{\partial z_j}{\partial w_{ji}}$$

$$= f'(z_j) w_{kj} \frac{\partial a_i w_{ji}}{\partial w_{ji}} = f'(z_j) w_{kj} a_i$$

put this result back into this for:

$$\frac{\partial C}{\partial w_{ji}} = \sum_k (a_k - t_k) f'(z_k)(f'(z_j) w_{kj} a_i)$$

$$= f'(z_j) a_i \sum_k (a_k - t_k) f'(z_k) w_{kj}$$

$$= f'(z_j) a_i \left(\sum w_{kj} \delta_k\right) = a_i \delta_j$$

$a_i = x_i$ in this case, as it is the input layer.

(the learning rate was forgotten here, but it's just an extra constant that is irrelevant to the derivation).

Task 1.1

BP1: assuming we know (proven in BP4, and in a specific case in assignment 1).

$$\alpha \frac{\partial C}{\partial w_{ji}} = \alpha \delta_j a_i$$

$$= \alpha \frac{\partial C}{\partial z_j} a_i$$

by chain rule:

$$= \alpha a_i \left(\sum_k \frac{\partial C}{\partial z_k} \frac{\partial z_k}{\partial z_j} \right) = \alpha a_i \left(\sum_k \frac{\partial z_k}{\partial z_j} \frac{\partial C}{\partial z_k} \right)$$

$$\overset{\alpha a_i}{=} \left(\sum_k \frac{\partial z_k}{\partial z_j} \delta_k \right) \overset{\alpha a_i}{=} \left(\sum \frac{\partial (w_{kj} \cdot a_j)}{\partial z_j} \delta_k \right.$$

$$\overset{\alpha a_i}{=} \sum_k \frac{\partial (w_{kj} \cdot f(z_j))}{\partial z_j} \delta_k$$

$$\overset{\alpha a_i}{=} f'(z_j) \sum_k w_{kj} \delta_k = \delta_j$$

$\Rightarrow$ which is as specified $\blacksquare$

($a_i = x_i$ again here, as it's the input).