Ryan Ott
14862565
ryan.ott@student.uva.nl

1 Linear Module

1.a

3. Substituting this into the expression for $\frac{\partial L}{\partial W}$:

$$\frac{\partial L}{\partial W} = \sum_{i,j} \frac{\partial L}{\partial Y_{ij}} \frac{\partial Y_{ij}}{\partial W_{mn}} = \sum_{i,j,k} \frac{\partial L}{\partial Y_{ij}} X_{ik} \delta_{km} \delta_{jn}$$

4. As the Kronecker deltas effectively filter for $k = m$ and $j = n$, the summation over k and j selects the m-th row and n-th column of X and $\frac{\partial L}{\partial Y}$ respectively. Thus:

$$\frac{\partial L}{\partial W_{mn}} = \sum_{i} X_{im} \frac{\partial L}{\partial Y_{in}}$$

5. Rewriting in matrix terms, we arrive at the final expression:

$$\frac{\partial L}{\partial W} = \left(\frac{\partial L}{\partial Y} \right)^T X$$

This correctly captures the gradients of the loss with respect to the weights W in a linear module, using matrix calculus and the chain rule.