Deep Learning Assignment 1

Jochem Loedeman
12995282

1 MLP backprop and NumPy Implementation

1.1 Evaluating the Gradients

Question 1.1

(a) Given that

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{W}^T + \boldsymbol{B},$$

we deduce that

$$Y_{nm} = \sum_p X_{mp} W_{np} + B_{nm}.$$

We will now find the required derivatives.

(i)

$$\frac{\partial L}{\partial W_{ij}} = \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} \frac{\partial Y_{nm}}{\partial W_{ij}}$$

But

$$\frac{\partial Y_{nm}}{\partial W_{ij}} = \sum_p X_{mp} \frac{\partial W_{np}}{\partial W_{ij}} = \sum_p X_{mp} \delta_{ni} \delta_{pj} = X_{mj} \delta_{ni},$$

and hence

$$\frac{\partial L}{\partial W_{ij}} = \sum_{m,n} \frac{\partial L}{\partial Y_{mn}} X_{mj} \delta_{ni} = \sum_m \frac{\partial L}{\partial Y_{mi}} X_{mj}.$$

In matrix-vector notation, this is equivalent to

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

(ii)