Why is $\frac{\partial J(W)}{\partial W_{i,j}^{(out)}} = \left(A^{(h)}\right)^T \delta^{(out)}$?

Kassi Bertrand

January 2023

1 Introduction

In this LaTeX document, I want to show my step-by-step process to derive the partial derivative of all the weights in $W^{(out)}$.

I'll start simple with a 2-2-2 MLP, which is a Multi-Layer Perceptron with 2 inputs, 2 hidden units, and 2 outputs, and then generalize what we learn to a general $m - d - t$ MLP.

In both cases, I will use the following loss/error function:

$$J(w) = -\sum_{i=1}^{n} \sum_{j=1}^{t} y_j^{[i]} ln(a_j^{[i]}) + (1 - y_j^{[i]}) ln(1 - a_j^{[i]})]$$

Ready? Let's go!

2 With a 2-2-2 MLP