

PadhAI: Backpropagation - the full version

One Fourth Labs

Quick recap of the story so far

What have we covered up till this point?

1. Our roadmap for this module
 - a. To calculate the desired gradient, we need to compute
 - b. Gradient w.r.t output units
 - c. Gradient w.r.t hidden units
 - d. Gradient w.r.t weights and biases
 - e.

$\frac{\partial L(\theta)}{\partial W_{111}} =$	$\frac{\partial L(\theta)}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial a_3}$	$\frac{\partial a_3}{\partial h_2} \frac{\partial h_2}{\partial a_2}$	$\frac{\partial a_2}{\partial h_1} \frac{\partial h_1}{\partial a_1}$	$\frac{\partial a_3}{\partial W_{111}}$
Talk to the weight directly	Talk to the output layer	Talk to the previous hidden layer	Talk to the previous hidden layer	Talk to the weights
		works for any number of output layers		

- f. For the rest of this exercise, our focus is on Cross Entropy loss and Softmax output.
2. Here, what the sections highlighted in green are what we have covered so far, i.e. the derivative with respect to the last layer a_L
3. The gradient was calculated to be $\nabla_{a_L} L(\theta) = \frac{\partial L(\theta)}{\partial a_{Li}} = -(y - \hat{y}_i)$