## **Quick recap of the story so far**

What have we covered up till this point?

1. Our roadmap for this module
   1. To calculate the desired gradient, we need to compute
   2. Gradient w.r.t output units
   3. Gradient w.r.t hidden units
   4. Gradient w.r.t weights and biases

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| 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 | |  |

* 1. For the rest of this exercise, our focus is on *Cross Entropy loss*  and *Softmax* output.

1. Here, what the sections highlighted in green are what we have covered so far, i.e. the derivative with respect to the last layer aL
2. The gradient was calculated to be