## **Computing derivatives w.r.t Hidden Layers**

### **Part 3**

1. Consider the next layer ai
   1. The first derivative is what we computed in part 2
   2. We need to compute the second derivative
   3. We know that hij is simply the application of an activation function (sigmoid, tanh etc) to aij
   4. So it can be rewritten as where hij = g(aij) and g’(aij) is its derivative
2. The full gradient can be written as

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  | |  |
| .  .  . | |
|  | |

* 1. This vector is the element-wise product of two vectors and [...,g’(aik),...] (which is a vector of derivations of the activation function w.r.t the pre-activation layer. They are both vectors of n-terms

1. Thus (⊙ refers to element-wise multiplication)
2. This formula can be applied to any of the hidden layers