## **Computing derivatives w.r.t Output Layer**

### **Part 1**

The first derivative in the chain

1. What we are actually interested in is:
   1. Where L = layer number, i = neuron (from 1 to k), l = index of correct output
   2. Here, we use the cross entropy loss function
   3. In the output layer L, assume we have neurons aL1, aL2 … aLk
   4. The output layer L involves applying the softmax function the all the neurons
   5. again, (l refers to the index of the correct output neuron)
   6. Thus, depends on all the neurons’ outputs as they all appear in the denominator, thereby making the derivative non-zero for all the output neurons
2. From the previous points, we know that depends on
3. The first part of the derivative is fairly straightforward (of the form )