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

### **Part 2**

1. Continuing from where we left off
2. Here, we know that (taking the l-th entry of the softmax function applied to vector aL)
3. where aL = [aL1, aL2 … aLk]
   1. Where
   2. Selecting the l-th entry would give us the value
   3. This is of the form
   4. Here
   5. Substitute the values and expand the formula
   6. Here, consider , this value is 0 for all values of i : 0 to k except for when i = *l*
   7. Thus, we use an indicator variable to denote that all other values except i=*l* resolve to 0
   8. Now consider , here i’ ranges from 1 to k. When taking the derivative, only the index i=i’ remains, which is simply a derivative of an exponent.
   9. This is can be rewritten in terms of the softmax function for the different variables
   10. We know that the Softmax function is ŷ, so we rewrite it.
4. After cancellation