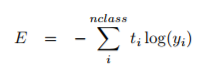
# Exercise 1 Report

The exercise was a good one to recap and renew old concepts of Neural Networks. The derivatives of the activation function were easy to compute and code. The code structure was very good and helpful. The bprop function in FullyConnectedLayer class took most of the time.

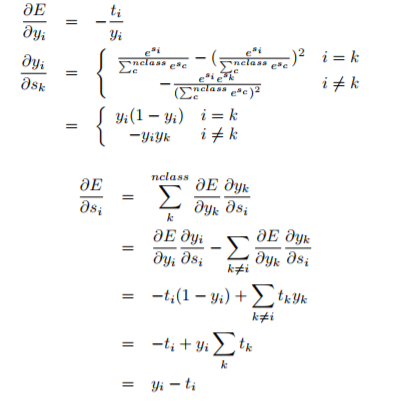
In the SoftmaxOutput class, the calculation of input\_grad was a complex procedure. It was computed as follows. The softmax activation of the ith output unit is



Thus, the error is



Therefore, computing the gradient yields the following result



Where is the predicted output