Part 1 Report : MNIST Dataset Accuracy

The model's accuracy has been increased to 96.3% by adding the following changes:

- Changing the optimizer to adam (which is a better optimizer in our case).

Adam optimizer inherits the good features of RMSProp and other algorithms. The results of the Adam optimizer are generally better than every other optimization algorithm, have faster computation time, and require fewer parameters for tuning.

- Adding more dense layers that are fully connected to each other.

In fact, each nonlinear activation function can be decomposed to Taylor series thus producing a polynomial of a degree higher than 1. By stacking several dense non-linear layers (one after the other) we can create higher and higher order of polynomials. For instance, let's imagine we use the following non-linear activation function: ($y=x^2+x$). By stacking 2 instances of it, we can generate a polynomial of degree 4, having (x^4, x^3, x^2, x) terms in it. Thus, the more layers we add, the more complex mathematical functions we can model.

- Adding the relu activation function for each dense layer (except the last one). Relu has shown an exponential increase in many models' accuracy, thus it is the best fit for MNIST dataset.

The main advantage of using the ReLU function over other activation functions is that it does not activate all the neurons at the same time. This means that the neurons will only be deactivated if the output of the linear transformation is less than 0.