

Omri Fridental

323869545

16.6.19

The Network built like that:

$n - layer_0(input) = 784 (28 * 28),$

$n - layer_1(hidden) = 84, (trial \text{ and } error)$

$n - layer_2(output) = 10,$

learning rate (eta) - by trial and error.

epochs = 20: afters many tryouts run, i concluded that the model converge with these parameters after around 10 epochs, so i gave spare 10.

initializing weights done like that: (known method)

$$\forall i,j : w_{ij}^{m * n} = normal(\mu = 0, \sigma = 1) * \sqrt{2 / n}$$

$$\forall i,j : b_{ij}^{m * n} = 0$$