Q1) we have inputted the correct variable for the gradient computation which are the loss and parameters. We need to minimize the loss between the predicted and the train value. As we have taken the gradient = tape.gradient(loss, parameters). Loss is the value that we are getting from the square of the train – predicted value.

Q2) We have neural networks of two layers which are 128 by 128 for each layer, the reason for having 100480 in the first layer is linear model which y = P0\*x + p1 where x is the input image which is 28 x 28 array equal to 728. So applying the above equation we can compute the value as, x\*p0 + p1 = 728 \* 128 + 128 which is equal to 100480.

Q3)

Following are the values:

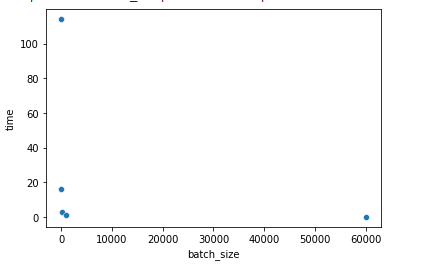
Batch\_size = 1: t=114s, accuracy= 0.8114

Batch\_size = 10: t=16s, accuracy= 0. 85070

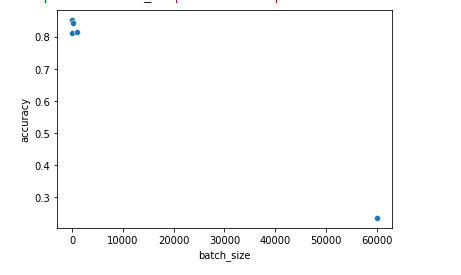
Batch\_size = 100: t=3s, accuracy= 0. 8429

Batch\_size = 100: t=1s, accuracy= 0. 81330

Batch\_size = 100: t=0s, accuracy= 0. 23589



Above graph is for the plot between batch size and the time taken to get trained.



Above graph is for the plot between batch size and accuracy to identify how much accuracy we are getting on any specific batch size.