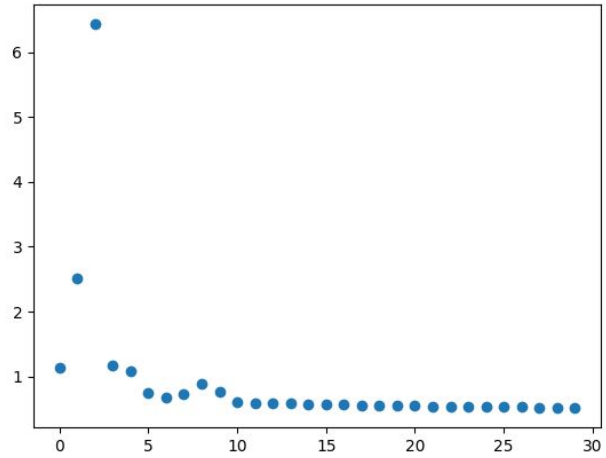
ASSIGNMENT-3

1.

Using Relu Activation

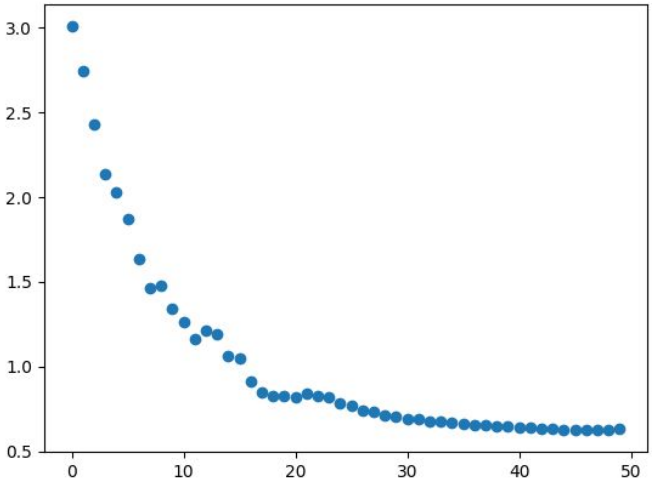
For subset:

Acc=0.9738



For entire database:

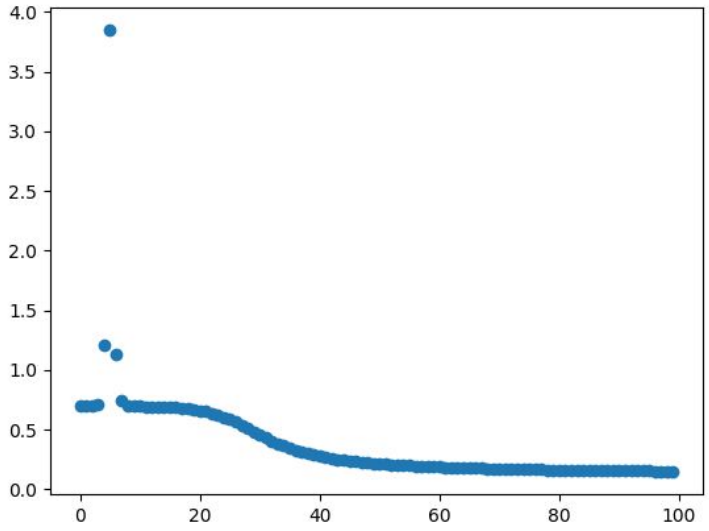
Acc=0.92



Using Sigmoid Activation

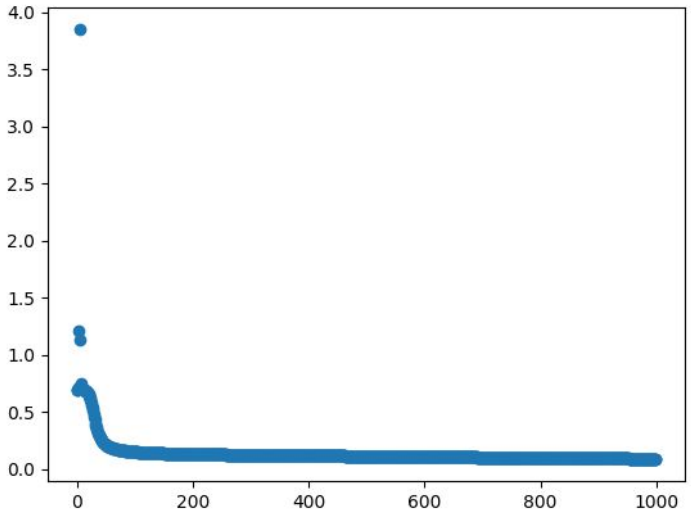
For subset:

Acc=0.83

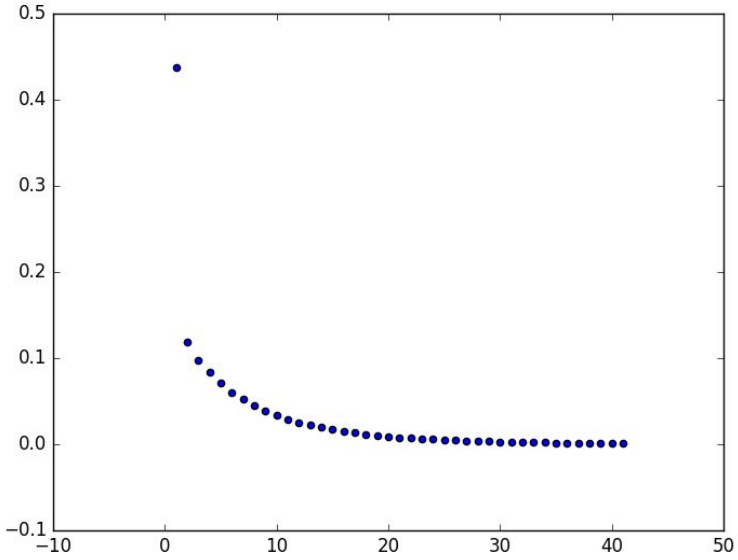


For entire database:

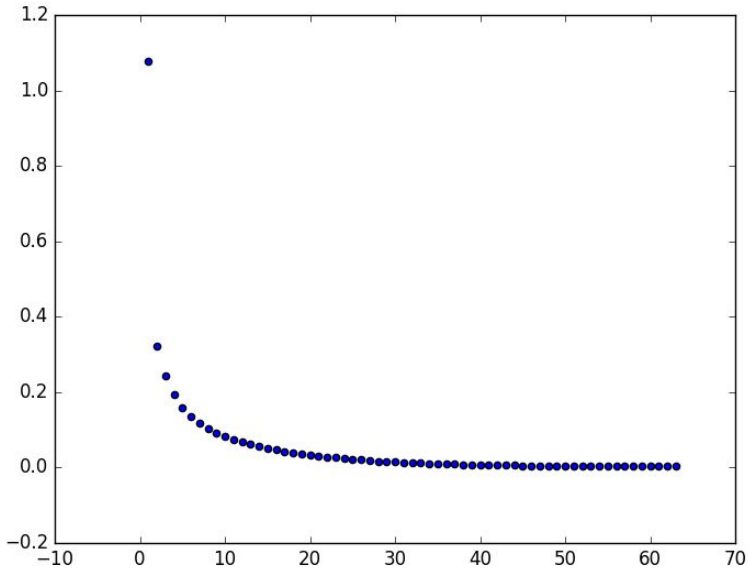
Acc=0.75



2. A) 0.983721 accuracy after 41 epochs
Diff in accuracy= 0.98731-0.9738=0.01351



B) 0.9783 accuracy after 63 epochs
Diff in accuracy=0.9873-0.92=0.0673



Values for 2a:

Iteration 1, loss = 0.43803080
Iteration 2, loss = 0.11922338
Iteration 3, loss = 0.09781124
Iteration 4, loss = 0.08375090
Iteration 5, loss = 0.07151051
Iteration 6, loss = 0.06039088
Iteration 7, loss = 0.05211490
Iteration 8, loss = 0.04443133
Iteration 9, loss = 0.03838058
Iteration 10, loss = 0.03344604
Iteration 11, loss = 0.02929022
Iteration 12, loss = 0.02538109
Iteration 13, loss = 0.02238846
Iteration 14, loss = 0.01986582
Iteration 15, loss = 0.01755445
Iteration 16, loss = 0.01547233
Iteration 17, loss = 0.01373389
Iteration 18, loss = 0.01185050
Iteration 19, loss = 0.01041231
Iteration 20, loss = 0.00916331
Iteration 21, loss = 0.00809742
Iteration 22, loss = 0.00731662
Iteration 23, loss = 0.00655079
Iteration 24, loss = 0.00572710
Iteration 25, loss = 0.00498084
Iteration 26, loss = 0.00445740
Iteration 27, loss = 0.00404931
Iteration 28, loss = 0.00359383
Iteration 29, loss = 0.00328862
Iteration 30, loss = 0.00290917
Iteration 31, loss = 0.00264382
Iteration 32, loss = 0.00237069
Iteration 33, loss = 0.00222710
Iteration 34, loss = 0.00198901
Iteration 35, loss = 0.00182203
Iteration 36, loss = 0.00169563
Iteration 37, loss = 0.00156758
Iteration 38, loss = 0.00146452
Iteration 39, loss = 0.00137392
Iteration 40, loss = 0.00128765

Iteration 41, loss = 0.00120624

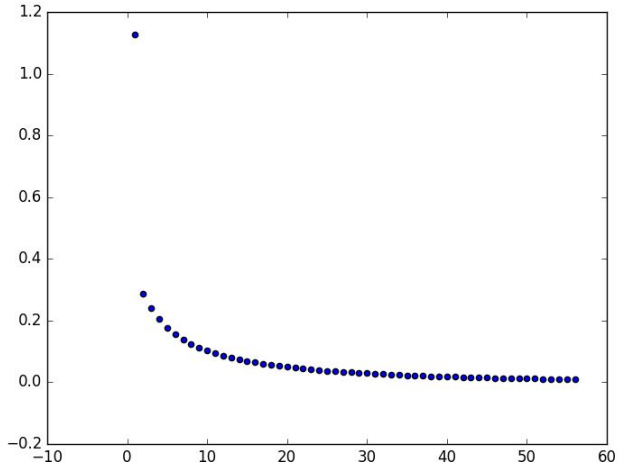
Values for 2b:

Iteration 1, loss = 1.07855417
Iteration 2, loss = 0.32247018
Iteration 3, loss = 0.24386968
Iteration 4, loss = 0.19342501
Iteration 5, loss = 0.15981504
Iteration 6, loss = 0.13577148
Iteration 7, loss = 0.11817061
Iteration 8, loss = 0.10376742
Iteration 9, loss = 0.09230813
Iteration 10, loss = 0.08305821
Iteration 11, loss = 0.07474951
Iteration 12, loss = 0.06810890
Iteration 13, loss = 0.06171708
Iteration 14, loss = 0.05587068
Iteration 15, loss = 0.05110689
Iteration 16, loss = 0.04714458
Iteration 17, loss = 0.04295232
Iteration 18, loss = 0.03965917
Iteration 19, loss = 0.03637980
Iteration 20, loss = 0.03392008
Iteration 21, loss = 0.03076510
Iteration 22, loss = 0.02812833
Iteration 23, loss = 0.02607275
Iteration 24, loss = 0.02418058
Iteration 25, loss = 0.02204834
Iteration 26, loss = 0.02036047
Iteration 27, loss = 0.01890287
Iteration 28, loss = 0.01714875
Iteration 29, loss = 0.01581448
Iteration 30, loss = 0.01459950
Iteration 31, loss = 0.01389374
Iteration 32, loss = 0.01282022
Iteration 33, loss = 0.01188058
Iteration 34, loss = 0.01090021
Iteration 35, loss = 0.01008262
Iteration 36, loss = 0.00962721
Iteration 37, loss = 0.00894215
Iteration 38, loss = 0.00835570
Iteration 39, loss = 0.00781065
Iteration 40, loss = 0.00734973

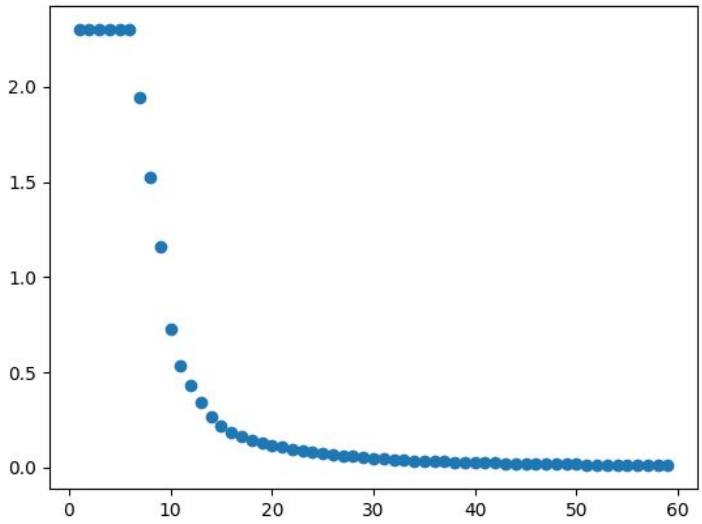
Iteration 41, loss = 0.00693390
Iteration 42, loss = 0.00645001
Iteration 43, loss = 0.00618357
Iteration 44, loss = 0.00587719
Iteration 45, loss = 0.00553606
Iteration 46, loss = 0.00523525
Iteration 47, loss = 0.00503540
Iteration 48, loss = 0.00484543
Iteration 49, loss = 0.00461258
Iteration 50, loss = 0.00436844
Iteration 51, loss = 0.00420978
Iteration 52, loss = 0.00404734
Iteration 53, loss = 0.00393061
Iteration 54, loss = 0.00378249
Iteration 55, loss = 0.00362651
Iteration 56, loss = 0.00353285
Iteration 57, loss = 0.00342026
Iteration 58, loss = 0.00333171
Iteration 59, loss = 0.00321654
Iteration 60, loss = 0.00310451
Iteration 61, loss = 0.00303361
Iteration 62, loss = 0.00297488
Iteration 63, loss = 0.00291925

---

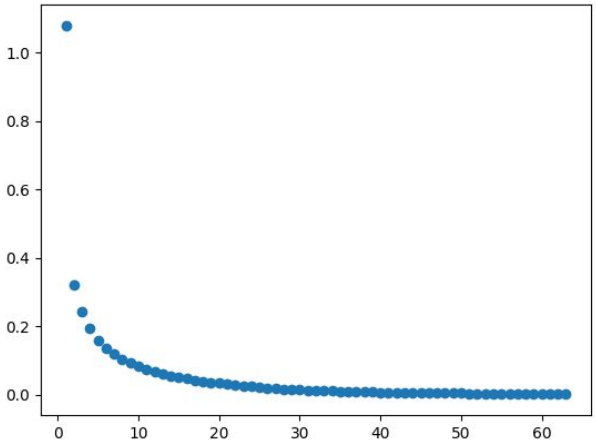
3. **Three models implemented were:**
    1. Simple network with one hidden layer of size 100; acc= 0.9768



    2. 4-layer network with sizes 100,50,25,5; acc= 0.9667



3. 2-layer network with sizes 100,75; acc=0.981



Best model:

Model 3 is best. Model 2 decreased acc, hence, we need to use 2 or 3 layer network. So, model 1 has underfitting and model 2 has overfitting. Model 3 with 2 layers is the ideal one.