Submitted by:
Raghav Sonavane
14EC35002

(Submitted codes are .py version of .ipynb)
Weights size is beyond the moodle permissible limit. They are uploaded on my github account URL:
https://github.com/raghavsonavane/Deep-Learning-CS60010

Although none of the .params file are required to run the codes, as I have attached necessary parameter files in the codes folder itself.

Task a

70:30 training:validation split
Input is 0-255 => 0-1 (without batch normalization)
Loss function: softmax cross entropy
Optimizer: sgd
Metric: Accuracy
Batch size=32
Initializer: Uniform
Learning rate=0.04

Vanilla Network 1

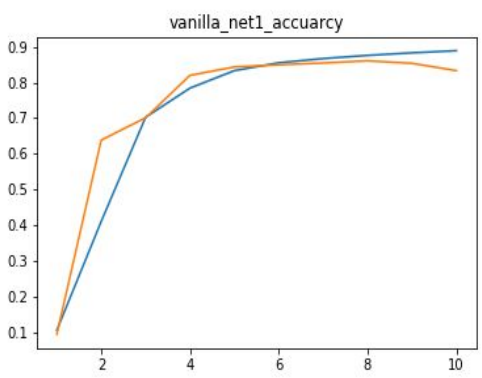
Epoch 0, training loss: 0.78, validation loss: 0.49
training accuracy: 0.7090870144706778, validation accuracy: 0.8317051509769094
Epoch 1, training loss: 0.48, validation loss: 0.45
training accuracy: 0.827756092916984, validation accuracy: 0.8400865896980462
Epoch 2, training loss: 0.41, validation loss: 0.40
training accuracy: 0.851223343488195, validation accuracy: 0.8559058614564832
Epoch 3, training loss: 0.38, validation loss: 0.37
training accuracy: 0.8618621477532369, validation accuracy: 0.8679507104795737
Epoch 4, training loss: 0.36, validation loss: 0.39
training accuracy: 0.869430693069307, validation accuracy: 0.8639542628774423

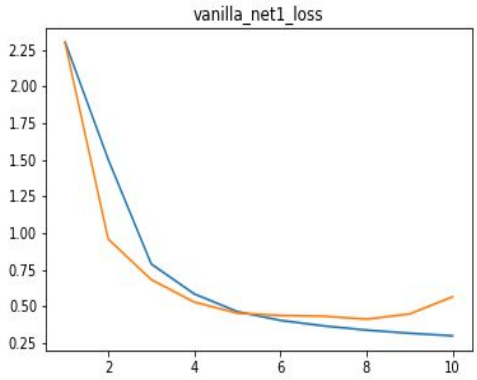
Final testing loss: 0.42
testing accuracy: 0.8564297124600639

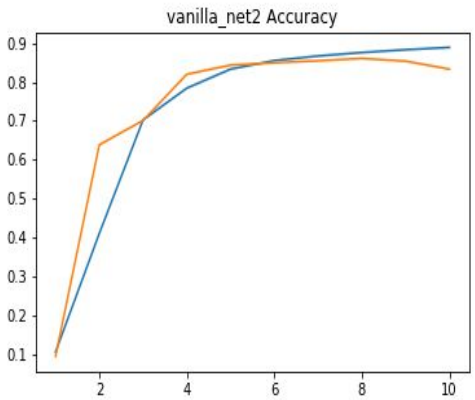
Vanilla Network 2

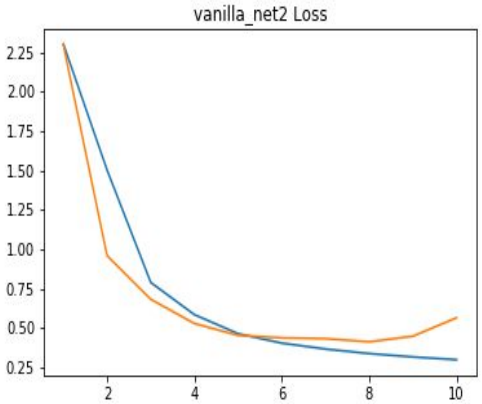
Epoch 0, training loss: 0.63, validation loss: 0.46
training accuracy: 0.7757044935262757, validation accuracy: 0.8284302841918295
Epoch 1, training loss: 0.42, validation loss: 0.39
training accuracy: 0.8503189261233816, validation accuracy: 0.8554618117229129
Epoch 2, training loss: 0.37, validation loss: 0.37
training accuracy: 0.8665984386900228, validation accuracy: 0.8648978685612788
Epoch 3, training loss: 0.33, validation loss: 0.36
training accuracy: 0.8771420411271896, validation accuracy: 0.8689498223801065
Epoch 4, training loss: 0.31, validation loss: 0.35
training accuracy: 0.8853055978674791, validation accuracy: 0.8733348134991119

Final testing loss: 0.36
testing accuracy: 0.8757987220447284









- Network 2 outperforms Network 2 due to its wider shape, which enables it to capture more information as compared to Network 1. Generally deeper networks perform better than wider ones, but the network should be able to capture information of the input data for better classification, which is signified in our task.
- As the data is not batch normalized, training is much dependent on the initializers. The result reported are the best ones, I got by running the code multiple times.

Task b

Network 2

Initializer: Xavier

Epoch 0, training loss: 0.58, validation loss: 0.43
training accuracy: 0.7978151180502666, validation accuracy: 0.8435279751332149
Epoch 1, training loss: 0.40, validation loss: 0.38
training accuracy: 0.8545792079207921, validation accuracy: 0.8620115452930728
Epoch 2, training loss: 0.35, validation loss: 0.36
training accuracy: 0.8696924980959635, validation accuracy: 0.8678952042628775
Epoch 3, training loss: 0.32, validation loss: 0.35
training accuracy: 0.8800694973343488, validation accuracy: 0.8715586145648313
Epoch 4, training loss: 0.30, validation loss: 0.34
training accuracy: 0.8871858339680122, validation accuracy: 0.8739453818827708
Epoch 5, training loss: 0.28, validation loss: 0.34
training accuracy: 0.8961586062452399, validation accuracy: 0.8752220248667851
Epoch 6, training loss: 0.26, validation loss: 0.34
training accuracy: 0.9028465346534653, validation accuracy: 0.8763876554174067
Epoch 7, training loss: 0.25, validation loss: 0.34
training accuracy: 0.909201256664128, validation accuracy: 0.8782193605683837
Epoch 8, training loss: 0.23, validation loss: 0.33
training accuracy: 0.915079969535415, validation accuracy: 0.8824378330373002
Epoch 9, training loss: 0.22, validation loss: 0.32
training accuracy: 0.9206730769230769, validation accuracy: 0.886101243339254

Final testing loss: 0.35
testing accuracy: 0.8801916932907349

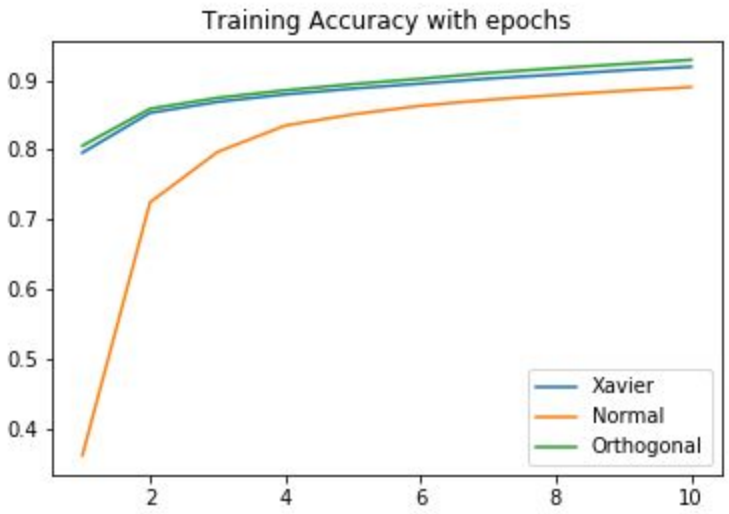
Initializer: Normal

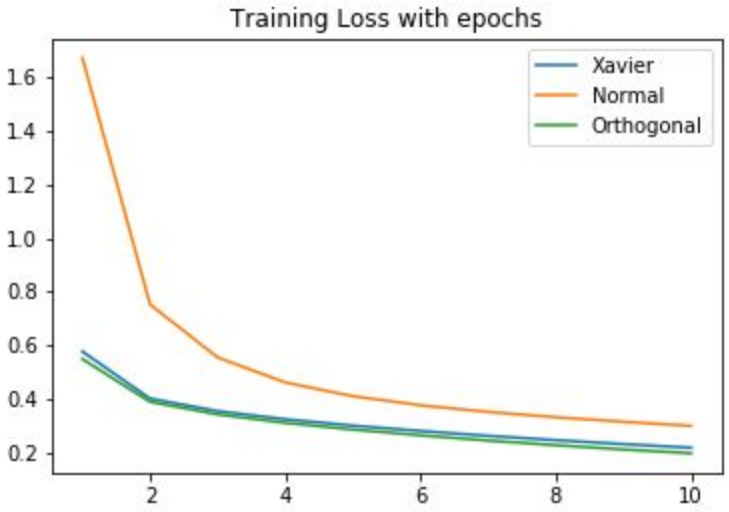
Epoch 0, training loss: 1.67, validation loss: 0.86
training accuracy: 0.3700971058644326, validation accuracy: 0.6572491119005328
Epoch 1, training loss: 0.76, validation loss: 0.65
training accuracy: 0.722534272658035, validation accuracy: 0.7508880994671403
Epoch 2, training loss: 0.55, validation loss: 0.51
training accuracy: 0.7991955445544554, validation accuracy: 0.82276865008881
Epoch 3, training loss: 0.46, validation loss: 0.44
training accuracy: 0.8341822162985529, validation accuracy: 0.8405861456483126
Epoch 4, training loss: 0.41, validation loss: 0.41
training accuracy: 0.851009139375476, validation accuracy: 0.8510768206039077
Epoch 5, training loss: 0.38, validation loss: 0.39
training accuracy: 0.8634091774562072, validation accuracy: 0.8595692717584369
Epoch 6, training loss: 0.35, validation loss: 0.38
training accuracy: 0.8729293602437167, validation accuracy: 0.8656749555950266
Epoch 7, training loss: 0.33, validation loss: 0.37
training accuracy: 0.8796648895658796, validation accuracy: 0.8701154529307282
Epoch 8, training loss: 0.31, validation loss: 0.36
training accuracy: 0.8859244097486672, validation accuracy: 0.8731682948490231
Epoch 9, training loss: 0.30, validation loss: 0.37
training accuracy: 0.891945925361767, validation accuracy: 0.8710590586145648
Final testing loss: 0.39
testing accuracy: 0.8623202875399361

Initializer: Orthogonal

Epoch 0, training loss: 0.55, validation loss: 0.42
training accuracy: 0.8058358720487433, validation accuracy: 0.8465253108348135

Epoch 1, training loss: 0.39, validation loss: 0.37
training accuracy: 0.8576256664127951, validation accuracy: 0.8631771758436945
Epoch 2, training loss: 0.34, validation loss: 0.36
training accuracy: 0.874785795887281, validation accuracy: 0.8692273534635879
Epoch 3, training loss: 0.31, validation loss: 0.34
training accuracy: 0.8847343869002284, validation accuracy: 0.8751110124333925
Epoch 4, training loss: 0.29, validation loss: 0.34
training accuracy: 0.8933263518659559, validation accuracy: 0.8776642984014209
Epoch 5, training loss: 0.27, validation loss: 0.33
training accuracy: 0.9006568926123382, validation accuracy: 0.88199378330373
Epoch 6, training loss: 0.25, validation loss: 0.33
training accuracy: 0.9083920411271896, validation accuracy: 0.8831039076376554
Epoch 7, training loss: 0.23, validation loss: 0.32
training accuracy: 0.9153417745620716, validation accuracy: 0.8865452930728241
Epoch 8, training loss: 0.22, validation loss: 0.32
training accuracy: 0.9216726961157654, validation accuracy: 0.8873223801065719
Epoch 9, training loss: 0.20, validation loss: 0.32
training accuracy: 0.927503808073115, validation accuracy: 0.8883769982238011
Final testing loss: 0.34
testing accuracy: 0.8812899361022364



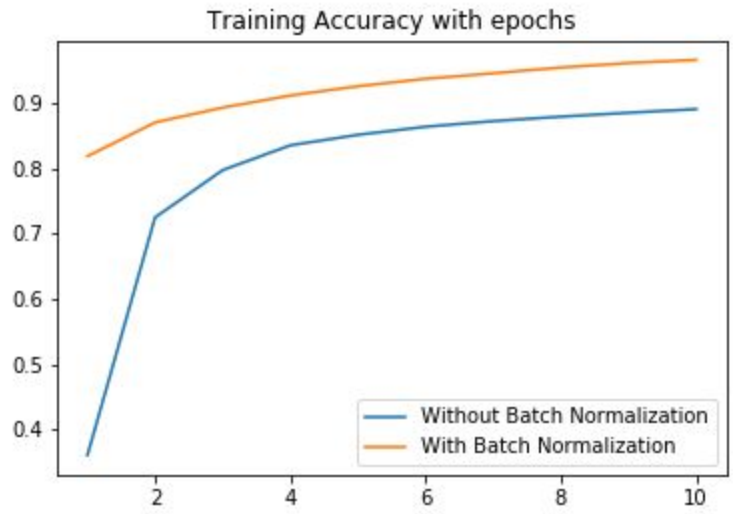


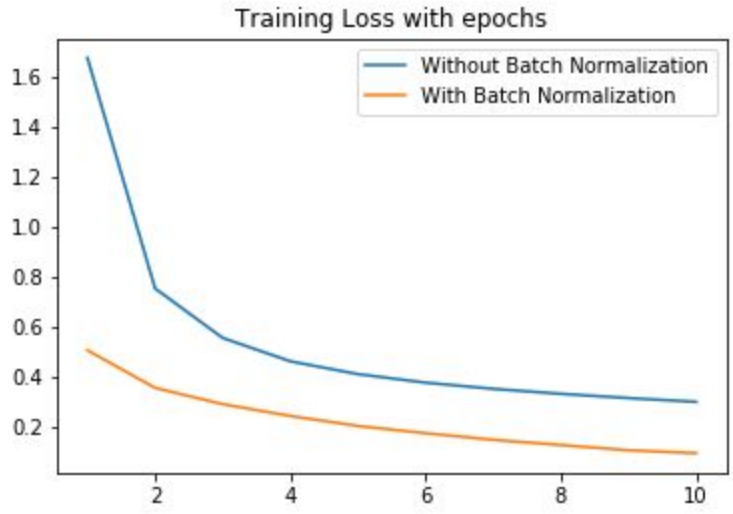
Experiment 2: batch normalization

- I have tried running my code with and without batch normalization both for (Uniform, Xavier, Normal and Orthogonal) initializers. Output loss and accuracy both are bad without batch normalization.
- Theoretically, without batch normalization as well the output should be same, but convergence rate becomes slower. It's can be seen that without batch normalization, the chances of getting stuck at local maxima is higher.
- I am reporting results for Normal initializer.

  Epoch 0, training loss: 0.51, validation loss: 0.37
  training accuracy: 0.8183311119573495, validation accuracy: 0.8625666074600356

Epoch 1, training loss: 0.36, validation loss: 0.35
training accuracy: 0.870049504950495, validation accuracy: 0.8743894316163411
Epoch 2, training loss: 0.29, validation loss: 0.35
training accuracy: 0.8927075399847677, validation accuracy: 0.8774422735346359
Epoch 3, training loss: 0.24, validation loss: 0.38
training accuracy: 0.911200495049505, validation accuracy: 0.8748889875666075
Epoch 4, training loss: 0.20, validation loss: 0.39
training accuracy: 0.9254093678598629, validation accuracy: 0.8772757548845471
Epoch 5, training loss: 0.17, validation loss: 0.43
training accuracy: 0.9367859862909368, validation accuracy: 0.8727797513321492
Epoch 6, training loss: 0.15, validation loss: 0.46
training accuracy: 0.9452113480578828, validation accuracy: 0.8726687388987566
Epoch 7, training loss: 0.13, validation loss: 0.49
training accuracy: 0.954017517136329, validation accuracy: 0.8693383658969804
Epoch 8, training loss: 0.11, validation loss: 0.49
training accuracy: 0.9608958492003047, validation accuracy: 0.8745004440497336
Epoch 9, training loss: 0.09, validation loss: 0.51
training accuracy: 0.965703541507997, validation accuracy: 0.8765541740674956
Final testing loss: 0.54
testing accuracy: 0.8699081469648562





Experiment 3: Dropout

Rate: 0.6

Epoch 0, training loss: 0.95, validation loss: 0.56
training accuracy: 0.6441355674028941, validation accuracy: 0.7910190941385435
Epoch 1, training loss: 0.63, validation loss: 0.47
training accuracy: 0.7718488194973343, validation accuracy: 0.8273756660746003
Epoch 2, training loss: 0.55, validation loss: 0.43
training accuracy: 0.8030512185833968, validation accuracy: 0.839642539964476
Epoch 3, training loss: 0.52, validation loss: 0.42
training accuracy: 0.81754569687738, validation accuracy: 0.8441940497335702
Epoch 4, training loss: 0.49, validation loss: 0.40
training accuracy: 0.8254950495049505, validation accuracy: 0.8524644760213144
Epoch 5, training loss: 0.47, validation loss: 0.39

training accuracy: 0.8336586062452399, validation accuracy: 0.8561278863232682
Epoch 6, training loss: 0.46, validation loss: 0.39
training accuracy: 0.8370620715917746, validation accuracy: 0.8565719360568383
Epoch 7, training loss: 0.44, validation loss: 0.38
training accuracy: 0.8417745620715917, validation accuracy: 0.861845026642984
Epoch 8, training loss: 0.43, validation loss: 0.36
training accuracy: 0.8462966488956588, validation accuracy: 0.8667295737122558
Epoch 9, training loss: 0.43, validation loss: 0.36
training accuracy: 0.8459158415841584, validation accuracy: 0.8689498223801065
Final testing loss: 0.38
testing accuracy: 0.8588258785942492

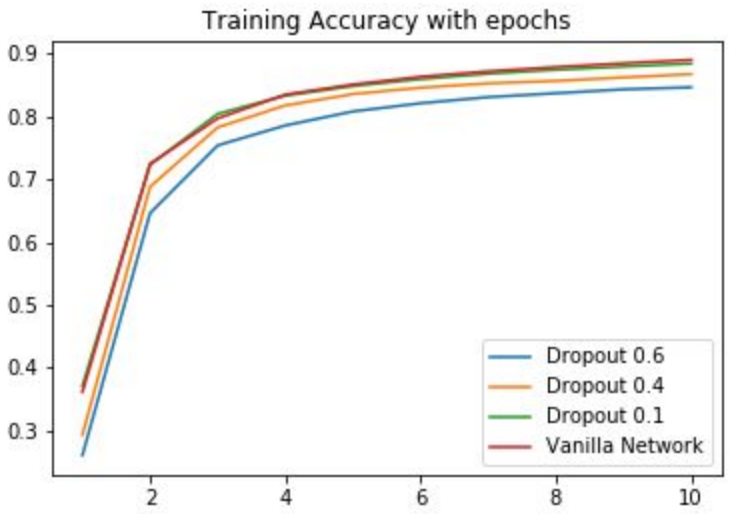
Rate: 0.4

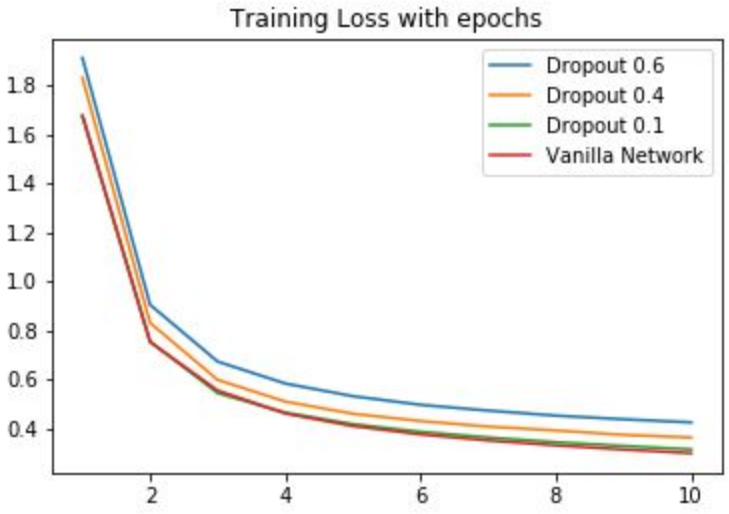
Epoch 0, training loss: 0.75, validation loss: 0.48
training accuracy: 0.7300552170601675, validation accuracy: 0.8148312611012434
Epoch 1, training loss: 0.51, validation loss: 0.41
training accuracy: 0.817759900990099, validation accuracy: 0.846636323268206
Epoch 2, training loss: 0.46, validation loss: 0.38
training accuracy: 0.8359434501142422, validation accuracy: 0.8599578152753108
Epoch 3, training loss: 0.43, validation loss: 0.38
training accuracy: 0.8474152703731912, validation accuracy: 0.8592917406749556
Epoch 4, training loss: 0.40, validation loss: 0.36
training accuracy: 0.85500761614623, validation accuracy: 0.8690608348134992
Epoch 5, training loss: 0.39, validation loss: 0.35
training accuracy: 0.8590774942878904, validation accuracy: 0.8713920959147424
Epoch 6, training loss: 0.37, validation loss: 0.36
training accuracy: 0.8648372048743336, validation accuracy: 0.8660079928952042
Epoch 7, training loss: 0.36, validation loss: 0.34
training accuracy: 0.8683596725057121, validation accuracy: 0.8767761989342806
Epoch 8, training loss: 0.35, validation loss: 0.34
training accuracy: 0.8719297410510282, validation accuracy: 0.8743339253996447
Epoch 9, training loss: 0.34, validation loss: 0.35
training accuracy: 0.8774752475247525, validation accuracy: 0.8736678507992895
Final testing loss: 0.37
testing accuracy: 0.8644169329073482

Rate: 0.1
Epoch 0, training loss: 0.61, validation loss: 0.43
training accuracy: 0.7844392612338157, validation accuracy: 0.8414742451154529
Epoch 1, training loss: 0.43, validation loss: 0.38
training accuracy: 0.8450352246763138, validation accuracy: 0.8604573712255773
Epoch 2, training loss: 0.38, validation loss: 0.36

training accuracy: 0.861005331302361, validation accuracy: 0.8669515985790408
Epoch 3, training loss: 0.35, validation loss: 0.35
training accuracy: 0.8706683168316832, validation accuracy: 0.8705595026642984
Epoch 4, training loss: 0.33, validation loss: 0.34
training accuracy: 0.8784986671744097, validation accuracy: 0.8757215808170515
Epoch 5, training loss: 0.31, validation loss: 0.33
training accuracy: 0.8849485910129474, validation accuracy: 0.879107460035524
Epoch 6, training loss: 0.29, validation loss: 0.32
training accuracy: 0.8904464965727342, validation accuracy: 0.8816052397868561
Epoch 7, training loss: 0.28, validation loss: 0.32
training accuracy: 0.895730198019802, validation accuracy: 0.8814942273534636
Epoch 8, training loss: 0.27, validation loss: 0.31
training accuracy: 0.8992764661081493, validation accuracy: 0.8879329484902309
Epoch 9, training loss: 0.26, validation loss: 0.31
training accuracy: 0.9046077684691546, validation accuracy: 0.8877664298401421
Final testing loss: 0.34
testing accuracy: 0.8793929712460063

- Dropout decreases the chances of overfitting the training data and also acts as a regulizer to give better testing performance. Accuracy/Loss curve is much smoother in this case and testing accuracy is also better.





Experiment 4: Optimizer comparison

SGD

Epoch 0, training loss: 0.72, validation loss: 0.50
training accuracy: 0.7642326732673267, validation accuracy: 0.8258769982238011
Epoch 1, training loss: 0.48, validation loss: 0.44
training accuracy: 0.8332063975628332, validation accuracy: 0.8439720248667851
Epoch 2, training loss: 0.43, validation loss: 0.41
training accuracy: 0.8505807311500381, validation accuracy: 0.8545182060390764
Epoch 3, training loss: 0.40, validation loss: 0.39
training accuracy: 0.8596963061690784, validation accuracy: 0.8609569271758437
Epoch 4, training loss: 0.37, validation loss: 0.38
training accuracy: 0.8667174409748667, validation accuracy: 0.8655639431616341

Epoch 5, training loss: 0.36, validation loss: 0.37
training accuracy: 0.8730245620715917, validation accuracy: 0.8702264653641207
Epoch 6, training loss: 0.34, validation loss: 0.36
training accuracy: 0.8779274562071592, validation accuracy: 0.8726132326820604
Epoch 7, training loss: 0.33, validation loss: 0.35
training accuracy: 0.8825209444021325, validation accuracy: 0.8751665186500888
Epoch 8, training loss: 0.32, validation loss: 0.34
training accuracy: 0.8861148134044173, validation accuracy: 0.8771647424511545
Epoch 9, training loss: 0.31, validation loss: 0.34
training accuracy: 0.8899466869763899, validation accuracy: 0.8795515097690941
Final testing loss: 0.36
testing accuracy: 0.8704073482428115

Adam
Epoch 0, training loss: 0.51, validation loss: 0.41
training accuracy: 0.8139042269611576, validation accuracy: 0.8461367673179396
Epoch 1, training loss: 0.38, validation loss: 0.39
training accuracy: 0.858482482863671, validation accuracy: 0.8542406749555951
Epoch 2, training loss: 0.34, validation loss: 0.36
training accuracy: 0.8752380045696877, validation accuracy: 0.8698379218472468
Epoch 3, training loss: 0.31, validation loss: 0.35
training accuracy: 0.8831635567402895, validation accuracy: 0.8741119005328597
Epoch 4, training loss: 0.29, validation loss: 0.36
training accuracy: 0.8904464965727342, validation accuracy: 0.8754440497335702
Epoch 5, training loss: 0.28, validation loss: 0.36
training accuracy: 0.8958015993907082, validation accuracy: 0.8767206927175843
Epoch 6, training loss: 0.26, validation loss: 0.38
training accuracy: 0.9019659177456207, validation accuracy: 0.8776087921847247
Epoch 7, training loss: 0.25, validation loss: 0.38
training accuracy: 0.9062023990860625, validation accuracy: 0.8812722024866785
Epoch 8, training loss: 0.23, validation loss: 0.39
training accuracy: 0.9093440594059405, validation accuracy: 0.8804951154529307
Epoch 9, training loss: 0.22, validation loss: 0.41
training accuracy: 0.9134615384615384, validation accuracy: 0.8813832149200711
Final testing loss: 0.48
testing accuracy: 0.8711062300319489

AdaDelta
Epoch 0, training loss: 0.66, validation loss: 0.46
training accuracy: 0.7581635567402895, validation accuracy: 0.8359791296625222
Epoch 1, training loss: 0.46, validation loss: 0.44
training accuracy: 0.8363242574257426, validation accuracy: 0.8391984902309059

Epoch 2, training loss: 0.41, validation loss: 0.38
training accuracy: 0.8519849581111958, validation accuracy: 0.8636767317939609
Epoch 3, training loss: 0.39, validation loss: 0.40
training accuracy: 0.860791127189642, validation accuracy: 0.8617895204262878
Epoch 4, training loss: 0.38, validation loss: 0.38
training accuracy: 0.8635043792840823, validation accuracy: 0.8689498223801065
Epoch 5, training loss: 0.37, validation loss: 0.44
training accuracy: 0.8687642802741813, validation accuracy: 0.866785079928952
Epoch 6, training loss: 0.36, validation loss: 0.50
training accuracy: 0.8706683168316832, validation accuracy: 0.8502997335701599
Epoch 7, training loss: 0.36, validation loss: 0.49
training accuracy: 0.8731197638994669, validation accuracy: 0.8662855239786856
Epoch 8, training loss: 0.36, validation loss: 0.43
training accuracy: 0.8751666031987814, validation accuracy: 0.8698379218472468
Epoch 9, training loss: 0.35, validation loss: 0.50
training accuracy: 0.8787604722010662, validation accuracy: 0.8705039964476021
Final testing loss: 0.53
testing accuracy: 0.8592252396166135

AdaGrad

Epoch 0, training loss: 0.52, validation loss: 0.38
training accuracy: 0.8168316831683168, validation accuracy: 0.8591807282415631
Epoch 1, training loss: 0.34, validation loss: 0.34
training accuracy: 0.8736909748667174, validation accuracy: 0.8733903197158082
Epoch 2, training loss: 0.30, validation loss: 0.32
training accuracy: 0.8886614623000761, validation accuracy: 0.8812722024866785
Epoch 3, training loss: 0.27, validation loss: 0.31
training accuracy: 0.8984910510281797, validation accuracy: 0.8845470692717584
Epoch 4, training loss: 0.25, validation loss: 0.31
training accuracy: 0.9075590251332826, validation accuracy: 0.8874888987566607
Epoch 5, training loss: 0.23, validation loss: 0.31
training accuracy: 0.914032749428789, validation accuracy: 0.8889320603907638
Epoch 6, training loss: 0.22, validation loss: 0.31
training accuracy: 0.9203398705255141, validation accuracy: 0.8885990230905861
Epoch 7, training loss: 0.20, validation loss: 0.31
training accuracy: 0.9247667555217061, validation accuracy: 0.8906527531083481
Epoch 8, training loss: 0.19, validation loss: 0.31
training accuracy: 0.9299552551408987, validation accuracy: 0.8917073712255773
Epoch 9, training loss: 0.18, validation loss: 0.31
training accuracy: 0.9342155369383092, validation accuracy: 0.8932615452930728
Final testing loss: 0.35
testing accuracy: 0.8835862619808307

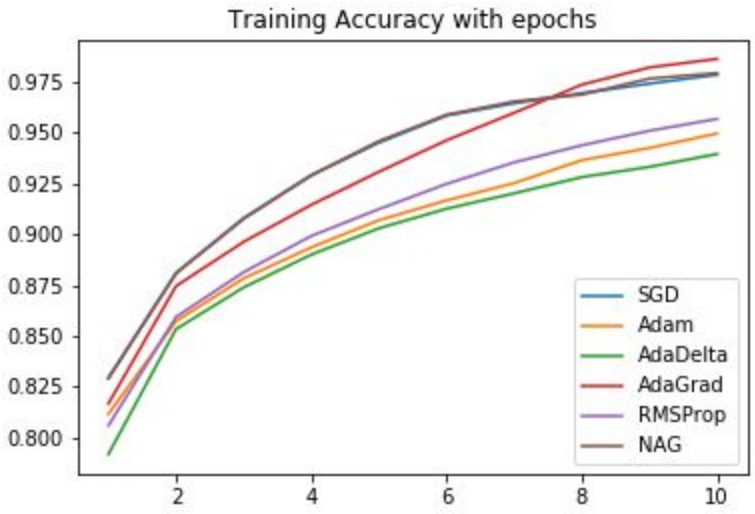
RMSProp

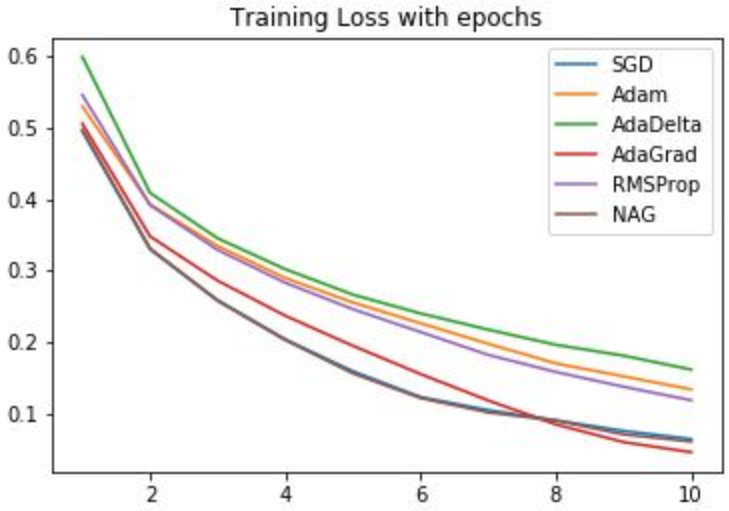
Epoch 0, training loss: 0.57, validation loss: 0.41
training accuracy: 0.7956968773800457, validation accuracy: 0.8568494671403197
Epoch 1, training loss: 0.44, validation loss: 0.39
training accuracy: 0.8480578827113481, validation accuracy: 0.865119893428064
Epoch 2, training loss: 0.41, validation loss: 0.41
training accuracy: 0.8589108910891089, validation accuracy: 0.8617340142095915
Epoch 3, training loss: 0.39, validation loss: 0.47
training accuracy: 0.8652418126428028, validation accuracy: 0.8594582593250444
Epoch 4, training loss: 0.38, validation loss: 0.44
training accuracy: 0.8703589108910891, validation accuracy: 0.8747224689165186
Epoch 5, training loss: 0.37, validation loss: 0.47
training accuracy: 0.8732149657273419, validation accuracy: 0.8652309058614565
Epoch 6, training loss: 0.37, validation loss: 0.53
training accuracy: 0.875523610053313, validation accuracy: 0.8584591474245116
Epoch 7, training loss: 0.37, validation loss: 0.47
training accuracy: 0.8754760091393755, validation accuracy: 0.8643428063943162
Epoch 8, training loss: 0.36, validation loss: 0.46
training accuracy: 0.878260662604722, validation accuracy: 0.8720581705150977
Epoch 9, training loss: 0.36, validation loss: 0.55
training accuracy: 0.8820449352627571, validation accuracy: 0.8723357015985791
Final testing loss: 0.63
testing accuracy: 0.865714856230032

Nesterov accelerated SGD

Epoch 0, training loss: 0.74, validation loss: 0.50
training accuracy: 0.7605436024371668, validation accuracy: 0.8264320603907638
Epoch 1, training loss: 0.48, validation loss: 0.44
training accuracy: 0.8328255902513328, validation accuracy: 0.844582593250444
Epoch 2, training loss: 0.43, validation loss: 0.41
training accuracy: 0.8485338918507236, validation accuracy: 0.8532415630550622
Epoch 3, training loss: 0.40, validation loss: 0.40
training accuracy: 0.858577684691546, validation accuracy: 0.8588476909413855
Epoch 4, training loss: 0.38, validation loss: 0.38
training accuracy: 0.8649324067022087, validation accuracy: 0.8638432504440497
Epoch 5, training loss: 0.36, validation loss: 0.37
training accuracy: 0.8710729246001523, validation accuracy: 0.868838809946714
Epoch 6, training loss: 0.34, validation loss: 0.36
training accuracy: 0.8762852246763138, validation accuracy: 0.8713920959147424
Epoch 7, training loss: 0.33, validation loss: 0.35
training accuracy: 0.8809501142421935, validation accuracy: 0.8733903197158082
Epoch 8, training loss: 0.32, validation loss: 0.34

training accuracy: 0.8849485910129474, validation accuracy: 0.8763321492007105
Epoch 9, training loss: 0.31, validation loss: 0.34
training accuracy: 0.8884234577303884, validation accuracy: 0.8783858792184724
Final testing loss: 0.36
testing accuracy: 0.8674121405750799





- Adagrad is much smoother and achieves best accuracy compared to rest of the optimizers, which can be because of the momentum (as it uses first and second gradient both) . Adadelta is not smooth, which might be because of the faster convergence to the optimal solution. Adam is combination of previous ones. RMSprop and ADAm both are

outperforming sgd as expected in this case. There is no rule of thumb in general, about which will perform better than rest. But ADAm and RMSprop are generally preferred.

Task c

Accuracy with original data => training: 0.8632 and testing: 0.8412
Accuracy with 1st hidden layer data => training: 0.9248 and testing: 0.884
Accuracy with 2nd hidden layer data => training: 0.9303 and testing: 0.8865
Accuracy with 3rd hidden layer data => training: 0.9246 and testing: 0.8895

- This results signifies that the hidden units represents the feature data in better way such that the classification task becomes easier. It can be concluded that the hidden units acts as a feature extractor from the raw data.