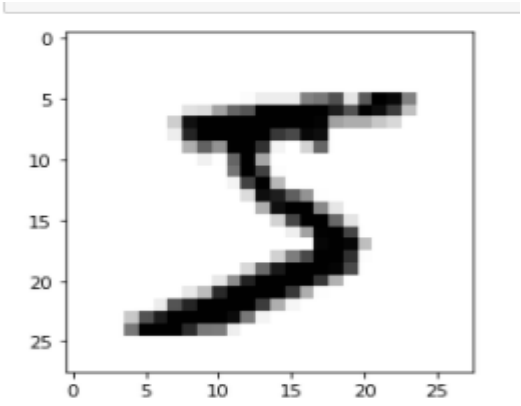
IT350 : Data Analytics
Lab Assignment 4

Name: Bhagyashri Bhamare **Roll No:181IT111**

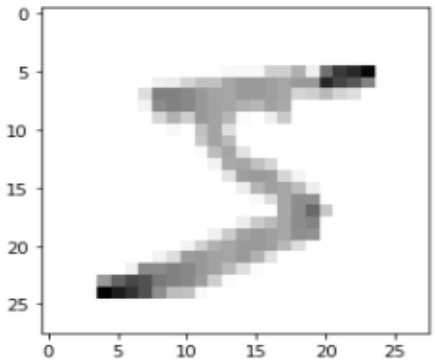
Mnist Dataset

Part -1

Data type-The MNIST database of handwritten digits, available from this page, has a training set of 60,000 examples, and a test set of 10,000 examples. It is a subset of a larger set available from NIST. The digits have been size-normalized and centered in a fixed-size image.



Trainset-



Accuracy-

```
Epoch 1/10
1875/1875 [==============================] - 4s 2ms/step - loss: 1.0145 - accuracy: 0.7371 - val_loss: 0.2611 - val_accuracy:
0.9221
Epoch 2/10
1875/1875 [==============================] - 4s 2ms/step - loss: 0.2472 - accuracy: 0.9276 - val_loss: 0.1934 - val_accuracy:
0.9407
Epoch 3/10
1875/1875 [==============================] - 4s 2ms/step - loss: 0.1766 - accuracy: 0.9487 - val_loss: 0.1537 - val_accuracy:
0.9531
Epoch 4/10
1875/1875 [==============================] - 3s 2ms/step - loss: 0.1327 - accuracy: 0.9604 - val_loss: 0.1262 - val_accuracy:
0.9614
Epoch 5/10
1875/1875 [==============================] - 3s 2ms/step - loss: 0.1081 - accuracy: 0.9685 - val_loss: 0.1130 - val_accuracy:
0.9655
Epoch 6/10
1875/1875 [==============================] - 4s 2ms/step - loss: 0.0878 - accuracy: 0.9740 - val_loss: 0.0970 - val_accuracy:
0.9705
Epoch 7/10
1875/1875 [==============================] - 4s 2ms/step - loss: 0.0739 - accuracy: 0.9782 - val_loss: 0.0932 - val_accuracy:
0.9713
Epoch 8/10
1875/1875 [==============================] - 4s 2ms/step - loss: 0.0607 - accuracy: 0.9823 - val_loss: 0.0873 - val_accuracy:
0.9729
Epoch 9/10
1875/1875 [==============================] - 4s 2ms/step - loss: 0.0526 - accuracy: 0.9848 - val_loss: 0.0822 - val_accuracy:
0.9733
Epoch 10/10
1875/1875 [==============================] - 4s 2ms/step - loss: 0.0431 - accuracy: 0.9878 - val_loss: 0.0819 - val_accuracy:
0.9734
Model: "sequential_12"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
flatten_12 (Flatten)         (None, 784)               0
_________________________________________________________________
dense_36 (Dense)             (None, 128)               100480
_________________________________________________________________
dense_37 (Dense)             (None, 64)                8256
_________________________________________________________________
dense_38 (Dense)             (None, 10)                650
=================================================================
Total params: 109,386
Trainable params: 109,386
Non-trainable params: 0
_________________________________________________________________
None
313/313 [==============================] - 0s 1ms/step - loss: 0.0819 - accuracy: 0.9734
Loss:  0.08191560953855515
Accuracy:  0.9733999967575073
```

Part-2

Dimension reduction using PCA

In [27]:
```
from sklearn.decomposition import PCA
```

In [28]:
```
pca = PCA(n_components=300)
```

In [ ]:
```
print(x_train.shape)
```

In [30]:
```
b=x_train[0].reshape(784)
```

In [31]:
```
print(b.shape)
```

```
(784,)
```

In [32]:
```
a=x_train.reshape(60000,784)
```

In [41]:
```
c=x_test.reshape(len(x_test),784)
```

In [33]:
```
print(a.shape)
```

```
(60000, 784)
```

In [35]:
```
print(b.shape)
```

```
(784,)
```

In [36]:
```
pca.fit(a)
```

Out[36]:
```
PCA(n_components=300)
```

In [38]:
```
transformed = pca.transform(a)
```

In [43]:
```
tr = pca.transform(c)
```

In [39]:
```
print(transformed.shape)
```

```
(60000, 300)
```

Accuracy after dimension reduction

```
Epoch 1/10
1875/1875 [==============================] - 2s 987us/step - loss: 1.3048 - accuracy: 0.6533 - val_loss: 0.3024 - val_accuracy:
0.9157
Epoch 2/10
1875/1875 [==============================] - 2s 1ms/step - loss: 0.2844 - accuracy: 0.9179 - val_loss: 0.2547 - val_accuracy:
0.9242
Epoch 3/10
1875/1875 [==============================] - 2s 1ms/step - loss: 0.2442 - accuracy: 0.9279 - val_loss: 0.2321 - val_accuracy:
0.9330
Epoch 4/10
1875/1875 [==============================] - 2s 1ms/step - loss: 0.2144 - accuracy: 0.9361 - val_loss: 0.2153 - val_accuracy:
0.9360
Epoch 5/10
1875/1875 [==============================] - 2s 1ms/step - loss: 0.1908 - accuracy: 0.9424 - val_loss: 0.2014 - val_accuracy:
0.9398
Epoch 6/10
1875/1875 [==============================] - 2s 1ms/step - loss: 0.1766 - accuracy: 0.9465 - val_loss: 0.1809 - val_accuracy:
0.9460
Epoch 7/10
1875/1875 [==============================] - 2s 1ms/step - loss: 0.1561 - accuracy: 0.9530 - val_loss: 0.1721 - val_accuracy:
0.9468
Epoch 8/10
1875/1875 [==============================] - 2s 1ms/step - loss: 0.1425 - accuracy: 0.9583 - val_loss: 0.1637 - val_accuracy:
0.9504
Epoch 9/10
1875/1875 [==============================] - 2s 1ms/step - loss: 0.1264 - accuracy: 0.9631 - val_loss: 0.1517 - val_accuracy:
0.9547
Epoch 10/10
1875/1875 [==============================] - 3s 1ms/step - loss: 0.1155 - accuracy: 0.9652 - val_loss: 0.1417 - val_accuracy:
0.9572
Model: "sequential_15"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_45 (Dense)             (None, 128)               38528
_________________________________________________________________
dense_46 (Dense)             (None, 64)                8256
_________________________________________________________________
dense_47 (Dense)             (None, 10)                650
=================================================================
Total params: 47,434
Trainable params: 47,434
Non-trainable params: 0
_________________________________________________________________
None
313/313 [==============================] - 0s 866us/step - loss: 0.1417 - accuracy: 0.9572
Loss:  0.1417023241519928
Accuracy:  0.9571999907493591
```

After dimension reduction to 300 accuracy changes from 97.33 to 95.71

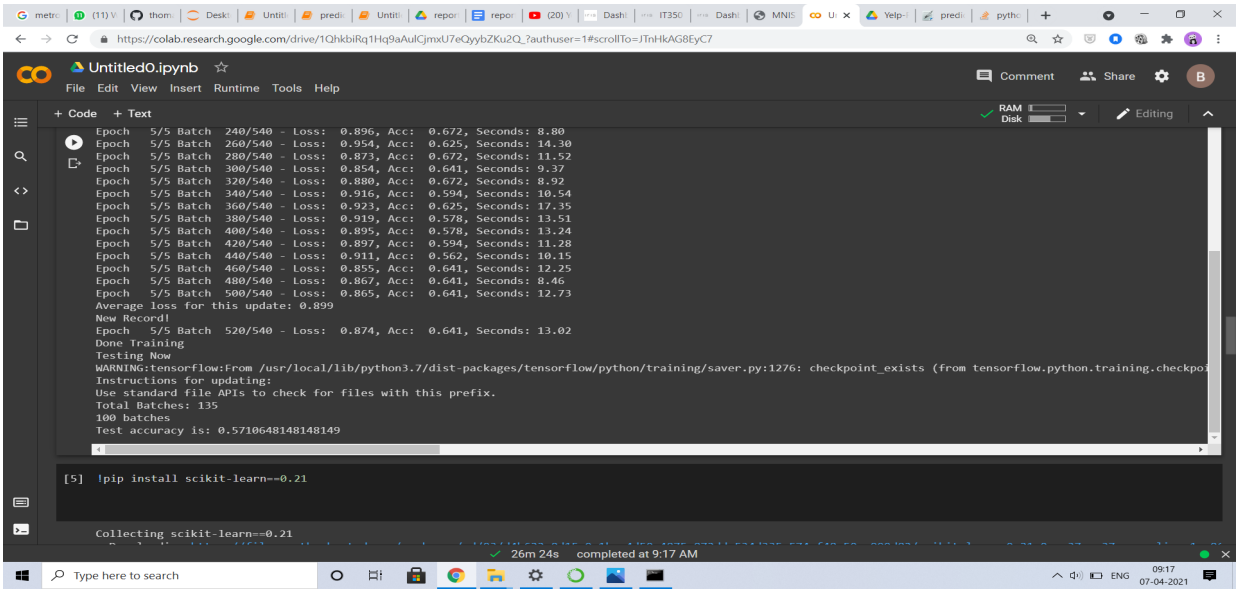
Accuracy for different Dimensions(linear)

dimensions	Accuracy
400	0.958000226020813
300	0.9572
200	0.9571999907493591
100	0.9571999907493591

Yelp dataset

Part -1

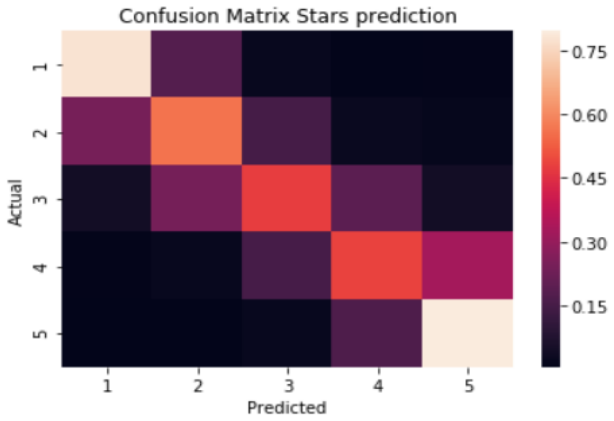
Accuarcy-



Testing

```
Please enter a review in englishI love this place! Will definitely come back!
INFO:tensorflow:Restoring parameters from ./saves/best_model.ckpt

You rated the restaurant: 5 stars!
```




```
<matplotlib.figure.Figure at 0x21949ce3f60>
```