

LAB 12

Try PCA with any classifier model.

Check the performance of classifier with and without PCA. [You are free to choose dataset and classifier of your choice]

Here I have taken SVM classifier with and without PCA.

```
In [2]: import tensorflow as tf
from tensorflow import keras
import matplotlib.pyplot as plt
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn import svm
from sklearn import metrics
from sklearn import decomposition
```

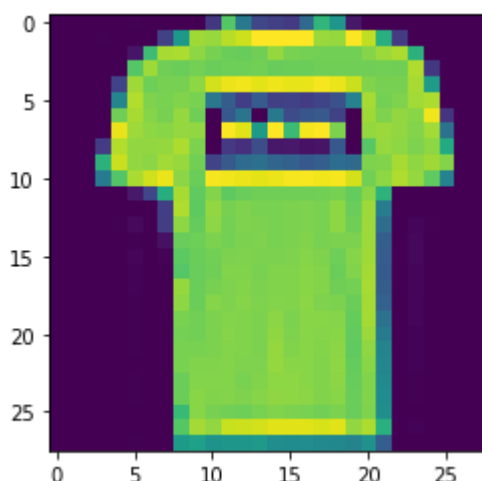
```
In [3]: fashion_mnist = keras.datasets.fashion_mnist
(X_train,y_train),(X_test,y_test) = fashion_mnist.load_data()
```

```
Downloading data from https://storage.googleapis.com/tensorflow/tf-keras-dat
assets/train-labels-idx1-ubyte.gz
32768/29515 [=====] - 0s 2us/step
Downloading data from https://storage.googleapis.com/tensorflow/tf-keras-dat
assets/train-images-idx3-ubyte.gz
26427392/26421880 [=====] - 23s 1us/step
Downloading data from https://storage.googleapis.com/tensorflow/tf-keras-dat
assets/t10k-labels-idx1-ubyte.gz
8192/5148 [=====] - 0s 1us/step
Downloading data from https://storage.googleapis.com/tensorflow/tf-keras-dat
assets/t10k-images-idx3-ubyte.gz
4423680/4422102 [=====] - 6s 1us/step
```

```
In [4]: class_names = ["T-shirt/top", "Trouser", "Pullover", "Dress", "Coat", "Sandal", "S
hirt", "Sneaker", "Bag", "Ankle"]
```

```
In [5]: plt.imshow(X_train[1])
```

```
Out[5]: <matplotlib.image.AxesImage at 0x7f91fccf3978>
```



```
In [6]: print("Label for above image is :",class_names[y_train[1]])
```

Label for above image is : T-shirt/top

```
In [7]: fig = plt.figure(figsize=(8, 6))
for i in range(15):
    ax = fig.add_subplot(3, 5, i + 1, xticks=[], yticks=[])
    ax.imshow(X_train[i], cmap=plt.cm.bone)
    ax.set_title(class_names[y_train[i]], fontsize='small', color='black')
```



```
In [8]: print("Image size is : " , X_train[1].shape)
```

Image size is : (28, 28)

```
In [9]: X_train, X_test = np.array(X_train, np.float32), np.array(X_test,np.float32)
num_features = 28*28
X_train, X_test = X_train.reshape([-1, num_features]), X_test.reshape([-1, num_features])
X_train, X_test = X_train/255, X_test/255
```

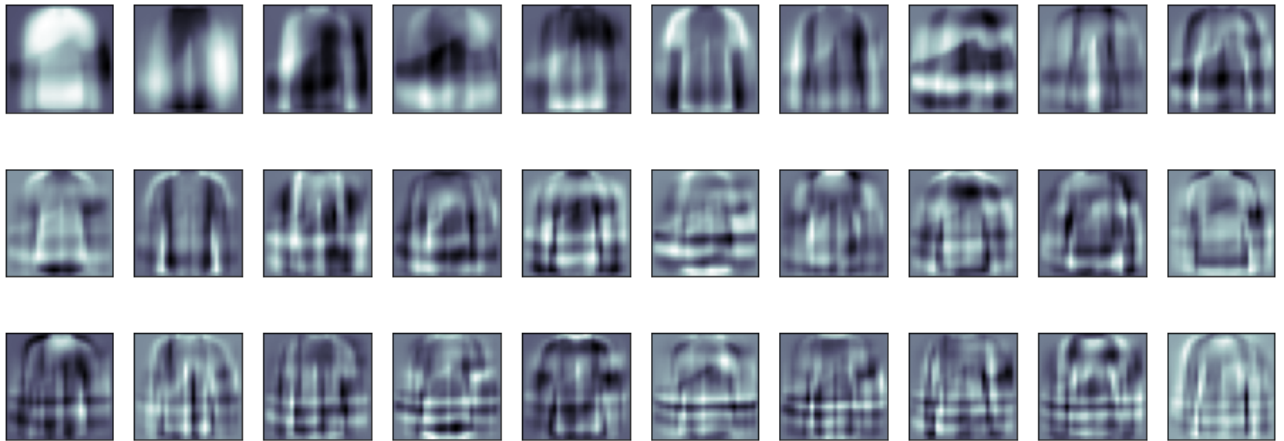
```
In [10]: pca = decomposition.PCA(n_components=400,whiten=True)
pca.fit(X_train)
```

Out[10]: PCA(n_components=400, whiten=True)

```
In [11]: print(pca.components_.shape)
```

(400, 784)

```
In [12]: fig = plt.figure(figsize=(16, 6))
for i in range(30):
    ax = fig.add_subplot(3, 10, i + 1, xticks=[], yticks=[])
    ax.imshow(pca.components_[i].reshape(28,28), cmap=plt.cm.bone)
```



```
In [13]: X_train_pca = pca.transform(X_train)
X_test_pca = pca.transform(X_test)
```

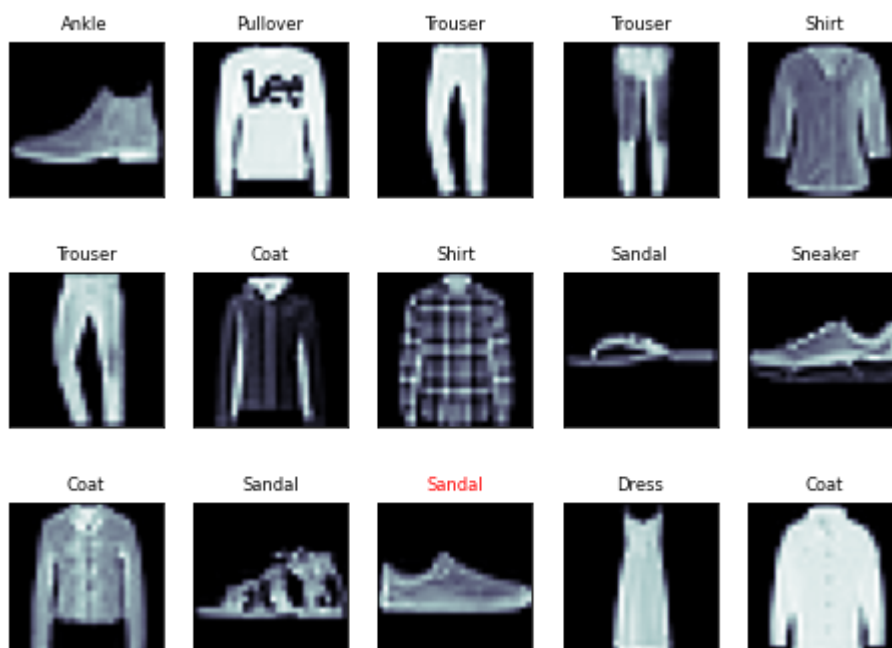
```
In [14]: print("X_train PCA shape : ",X_train_pca.shape)
print("X_test PCA shape : ",X_test_pca.shape)
```

```
X_train PCA shape : (60000, 400)
X_test PCA shape : (10000, 400)
```

```
In [15]: clf = svm.SVC(C=5., gamma=0.001, kernel='rbf')
clf.fit(X_train_pca, y_train)
```

```
Out[15]: SVC(C=5.0, gamma=0.001)
```

```
In [16]: fig = plt.figure(figsize=(8, 6))
for i in range(15):
    ax = fig.add_subplot(3, 5, i + 1, xticks=[], yticks=[])
    ax.imshow(X_test[i].reshape(28,28), cmap=plt.cm.bone)
    y_pred = clf.predict(X_test_pca[i, np.newaxis])[0]
    color = ('black' if y_pred == y_test[i] else 'red')
    ax.set_title(class_names[y_pred], fontsize='small', color=color)
```



```
In [17]: y_pred = clf.predict(X_test_pca)
print(metrics.classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.82	0.86	0.84	1000
1	0.99	0.97	0.98	1000
2	0.81	0.81	0.81	1000
3	0.88	0.90	0.89	1000
4	0.82	0.83	0.82	1000
5	0.96	0.97	0.97	1000
6	0.73	0.66	0.69	1000
7	0.95	0.97	0.96	1000
8	0.97	0.97	0.97	1000
9	0.98	0.96	0.97	1000
accuracy			0.89	10000
macro avg	0.89	0.89	0.89	10000
weighted avg	0.89	0.89	0.89	10000

```
In [18]: print("Accuracy : ",metrics.accuracy_score(y_test,y_pred))

Accuracy : 0.8902
```

Performance of classifier without PCA

```
In [19]: clf = svm.SVC(C=5., gamma=0.001, kernel='rbf')
clf.fit(X_train, y_train)
```

```
Out[19]: SVC(C=5.0, gamma=0.001)
```

```
In [20]: y_pred = clf.predict(X_test)
print(metrics.classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.81	0.84	0.82	1000
1	0.98	0.96	0.97	1000
2	0.76	0.78	0.77	1000
3	0.85	0.88	0.87	1000
4	0.77	0.79	0.78	1000
5	0.95	0.93	0.94	1000
6	0.68	0.59	0.63	1000
7	0.91	0.94	0.92	1000
8	0.95	0.97	0.96	1000
9	0.95	0.94	0.94	1000
accuracy			0.86	10000
macro avg	0.86	0.86	0.86	10000
weighted avg	0.86	0.86	0.86	10000

```
In [21]: print("Accuracy : ",metrics.accuracy_score(y_test,y_pred))

Accuracy : 0.8616
```

Note : SVM classifier with PCA performs model with accuracy 0.8902 and without PCA performs with accuracy 0.8616. Hence, For long range of feature set PCA might be helpful.