

## ✓ ML EX 8

Download MNIST dataset, apply PCA from scratch.

```
import numpy as np
import pandas as pd
import scipy.linalg
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df = pd.read_csv("mnist.csv")
```

```
label = df.label.astype(int)
df.drop("label", axis=1, inplace=True)
df.shape
```

```
(42000, 784)
```

```
standardized_data = StandardScaler().fit_transform(df)
```

```
sample_data = standardized_data
cov_matrix = np.matmul(sample_data.T, sample_data)
cov_matrix.shape
```

```
(784, 784)
```

```
eig_values, eig_vectors = eigh(cov_matrix)
print(eig_values)
print(eig_vectors)
```

```
[-5.15488736e-10 -3.04556546e-10 -1.99584487e-10 -1.67903046e-10
-8.62024010e-11 -5.35285609e-11 -2.54177119e-11 -1.06818692e-11
-6.40800111e-12 -5.32494636e-12 -2.38804230e-12 -2.07503739e-12
-1.86793799e-12 -1.19219889e-12 -8.32078263e-13 -6.07662011e-13
-4.29829311e-13 -1.14297629e-13 -8.01164397e-14 -6.79498373e-16
-7.83598132e-29 -6.75714114e-29 -4.85498540e-29 -7.68169369e-30
0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
5.84602126e-17 1.36134284e-15 2.23378206e-15 4.50347950e-15
4.58204919e-15 1.21363896e-14 1.85432157e-13 2.00602881e-13
2.77197133e-13 2.99924193e-13 3.75878409e-13 4.40152038e-13
4.49007310e-13 5.45926663e-13 5.46480173e-13 6.65366099e-13
1.09025464e-12 1.41355952e-12 2.24963238e-12 3.32477582e-12
4.32782718e-12 7.07349229e-12 8.71013600e-12 8.94074762e-12
9.25160139e-12 9.49723169e-12 1.13136566e-11 1.37713182e-11
2.04943155e-11 3.40000027e-11 3.74203885e-11 5.64570093e-11
8.49776462e-11 9.10268735e-11 1.23137743e-10 1.41363108e-10
1.79494753e-10 2.08398593e-10 2.48883534e-10 6.29292833e-10
2.50869230e+01 7.89904719e+01 1.99652010e+02 3.55210775e+02
4.93758275e+02 7.57430445e+02 7.79064240e+02 8.45892475e+02
8.72648025e+02 8.88753204e+02 8.99062941e+02 9.25014723e+02
9.57892904e+02 9.72704333e+02 9.80504167e+02 9.85361685e+02
9.98852435e+02 1.01031611e+03 1.01869487e+03 1.02844143e+03
1.04218562e+03 1.06115786e+03 1.06543862e+03 1.07809421e+03
1.09235135e+03 1.10830365e+03 1.11512028e+03 1.12578435e+03
1.12938332e+03 1.13747427e+03 1.15073485e+03 1.16136203e+03
1.16689869e+03 1.18255189e+03 1.19481789e+03 1.20234494e+03
1.20939684e+03 1.22395354e+03 1.23065511e+03 1.24075404e+03
1.25293225e+03 1.27149098e+03 1.28299954e+03 1.29007193e+03
1.29888692e+03 1.31582286e+03 1.32444292e+03 1.33328500e+03
1.33761938e+03 1.34512131e+03 1.35952549e+03 1.36938248e+03
1.38026016e+03 1.38632221e+03 1.39391639e+03 1.41793042e+03
1.43393005e+03 1.43909384e+03 1.44581040e+03 1.45619065e+03
1.46237983e+03 1.47375590e+03 1.48522466e+03 1.49380796e+03
1.50341604e+03 1.52181926e+03 1.53073604e+03 1.54321311e+03
1.55019476e+03 1.56082354e+03 1.57027786e+03 1.57084826e+03
1.58689403e+03 1.59151239e+03 1.61013355e+03 1.61036163e+03
1.62727207e+03 1.64858831e+03 1.65663570e+03 1.66145800e+03
1.67011803e+03 1.68073656e+03 1.68771926e+03 1.69021361e+03
1.69799094e+03 1.71198344e+03 1.72829267e+03 1.73522587e+03
1.75041137e+03 1.76451000e+03 1.78486476e+03 1.79067111e+03
1.79449832e+03 1.81100580e+03 1.82319665e+03 1.83117804e+03
1.83682350e+03 1.85191123e+03 1.85864558e+03 1.86911940e+03
1.89007681e+03 1.89495977e+03 1.90016750e+03 1.90963542e+03
1.91673692e+03 1.93540419e+03 1.95216753e+03 1.96211931e+03
1.97251372e+03 1.98313393e+03 1.99725757e+03 1.99898142e+03
```

```

2.01132346e+03 2.01876596e+03 2.03470291e+03 2.04745441e+03
2.06436135e+03 2.06899594e+03 2.07175828e+03 2.08646794e+03
2.10488303e+03 2.10980005e+03 2.13210651e+03 2.14680338e+03
2.15422115e+03 2.15997094e+03 2.18086019e+03 2.19054852e+03
2.19124451e+03 2.21096317e+03 2.22262530e+03 2.23793980e+03
2.24928106e+03 2.25108139e+03 2.27742818e+03 2.30551893e+03
2.31502958e+03 2.31574398e+03 2.32845663e+03 2.34048103e+03
2.34573119e+03 2.35404520e+03 2.36934359e+03 2.39609105e+03
2.40723041e+03 2.42030509e+03 2.43392711e+03 2.45041162e+03
~ ~ ~ ~

```

```

eig_values = eig_values[-2:]
eig_vectors = eig_vectors[:, -2:]
eig_vectors = eig_vectors.T

```

```

reduced_data = np.matmul(eig_vectors, sample_data.T)
reduced_data.shape

```

```
(2, 42000)
```

```

reduced_data = np.vstack((reduced_data, label))
reduced_data = reduced_data.T

```

```

reduced_df = pd.DataFrame(reduced_data, columns=['X', 'Y', 'label'])
reduced_df.label = reduced_df.label.astype(int)
reduced_df.head()

```

	X	Y	label
0	-5.226445	-5.140478	1
1	6.032996	19.292332	0
2	-1.705813	-7.644503	1
3	5.836139	-0.474207	4
4	6.024818	26.559574	0

```
g = sns.FacetGrid(reduced_df, hue='label').map(plt.scatter, 'X', 'Y').add_legend()
```

