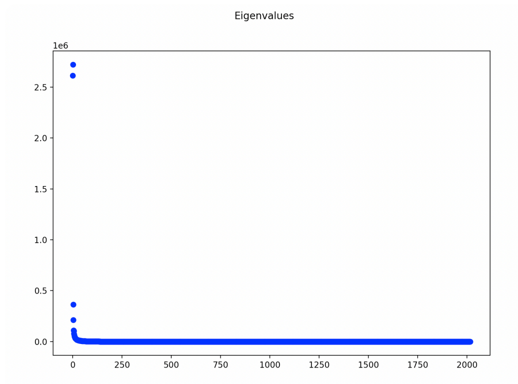
Berkay BARLAS
MTK 03734294
Introduction to ML

Problem 1

1.1)



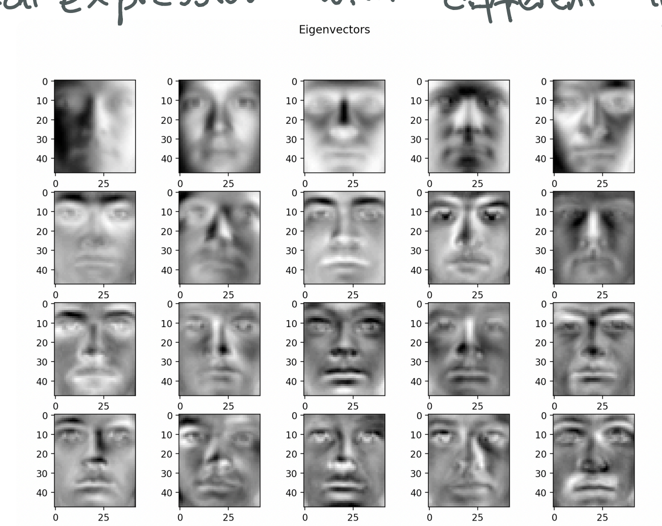

```
➜ Homework-11 git:(master) ✗ python3 calculations.py
42 principal components are needed for representing %95 of the total variation
166 principal components are needed for representing %99 of the total variation
```

For %95, reduction is $\frac{2016-42}{2016} \times 100 = 97.91\%$

For %99, reduction is $\frac{2016-166}{2016} \times 100 = 91.76\%$

1.2)

The first 20 eigenface have neutral facial expression with different lightning variations.



```python
import numpy as np
import matplotlib.pyplot as plt

yalefaces= np.loadtxt('yalefaces.csv', delimiter=',')

covarience = np.cov(yalefaces)
eigenVal, eigenVec = np.linalg.eig(covarience)
foundTotal = False
variances = []

sumOfVariances = 0

# Find Variances
for val in eigenVal:
    var = val / np.sum(eigenVal)
    variances.append(var)
# Find Principal Components
for i in range(len(variances)):
    sumOfVariances += variances[i]
    if sumOfVariances >= 0.95 and not foundTotal:
        print(*[ i, 'principal components are needed for representing %95 of the total variation'])
        foundTotal = True
    if sumOfVariances >= 0.99:
        print(*[ i, 'principal components are needed for representing %99 of the total variation'])
        break

fig = plt.figure()
X = np.linspace(0,len(eigenVal),len(eigenVal))
fig.suptitle('Eigenvalues')
plt.scatter(X, eigenVal, color='blue')
plt.show()

fig = plt.figure()
fig.suptitle('Eigenvectors')
for k in range(0,20):
    plt.subplot(4, 5, k+1)
    plt.imshow(eigenVec[:,k].reshape((48,42)),cmap='gray')
plt.show()
```

Problem 2

2.1) $\{ x : x = Vz + \mu, z \in \mathbb{R}^k \}$

If mean was zero

$$\frac{1}{n} \sum_{i=1}^{n} (\langle x_i, v \rangle)^2 = \frac{1}{n} \| Xv \|_2^2 = \frac{1}{n} v^T X^T X v$$

$$\sum_{i=1}^{n} (\langle x_i, v \rangle)^2 = \| Xv \|_2^2 = v^T X^T X v$$

V_{z_i} is maximized with largest eigenvalue of matrix $X^T X$

thus, $\min_{\mu, V, \{z_i\}} \sum_{i=1}^{n} \| x_i - V z_i \|_2^2 = n \sum_{j=k+1}^{d} \lambda_j$

with $\mu \in \mathbb{R}^d$ mean

$$\min_{\mu, V, \{z_i\}} \sum_{i=1}^{n} \| x_i - \mu - V z_i \|_2^2 = n \sum_{j=k+1}^{d} \lambda_j$$

where V ranges over all $d \times k$ matrices with orthonormal columns and λ_j : is jth largest eigenvalue of the emprical covariance $S = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T$