3 Task 3

The code of principal component analysis:

```
def zeroMean(dataMat):
    meanVal=np.mean(dataMat, axis=0)
    newData=dataMat-meanVal
    return newData,meanVal

def pca(dataMat,n):
    newData,meanVal=zeroMean(dataMat)
    covMat=np.cov(newData, rowvar=0)

    eigvals,eigVects=np.linalg.eig(np.mat(covMat))
    eigValIndice=np.argsort(eigvals)
    # print(eigValIndice)
    n_eigvalIndice=eigValIndice[ -1:-(n+1):-1] # pick n values
       from last to start
    # print(n_eigvalIndice)

    # choose the smallest one
    # n_eigvalIndice=1  # 1 is the index of the smallest
       eigenvalue

    n_eigvect=eigVects[ :,n_eigvalIndice]
    # print(n_eigvect)
    lowDDataMat=newData*n_eigvect
    reconMat=( lowDDataMat*n_eigvect.T)+meanVal
    return eigvals,eigVects,lowDDataMat ,reconMat
```

3.1 a)

The result of eigenvalue decomposition of the corresponding sample covariance matrix is [1.5 , 0.16666667]

3.2 b

1D representation under the**largest** eigenvalue is [[**-1.06066017**],[**-1.06066017**],[**1.06066017**],[**1.06066017**]]. And 1D representation under the **smallest** eigenvalue is [[**-0.35355339**],[**0.35355339**],[**-0.35355339**],[**0.35355339**]]

3.3 c

The code of disance computation

```
def get_dis_pairs(M):
    lens=M.shape[0]
    result=np.zeros((lens,lens))
    for i in range(lens):
        j=i+1
        while(j<lens):
            result[i][j]=np.linalg.norm(M[i]-M[j])
            j=j+1
    return result
```

After euclidean distance computation, we can see the result from Fig.2. The distance between points in original matrix and reconstructed matrix after PCA are the same. But, the distance between points after PCA changed due to the dimension decreasing.

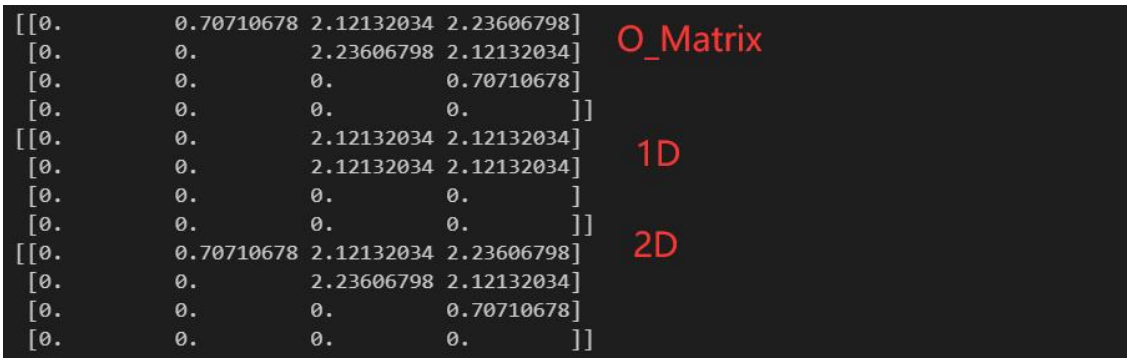
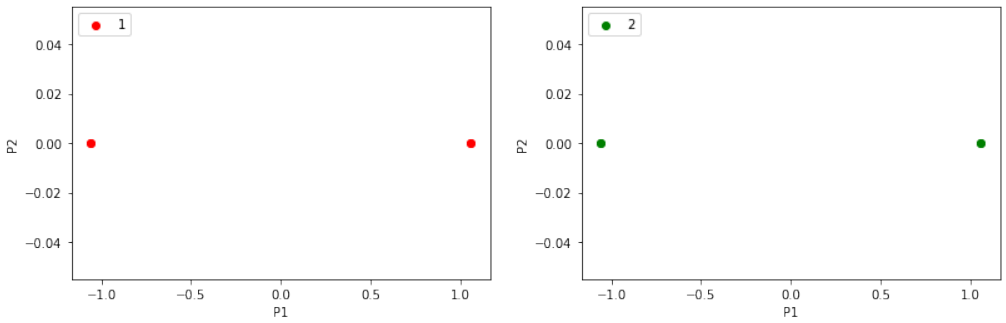


Figure 2: Euclidean distance

3.4 d)

In new dataset. the eigenvalue decomposition of the corresponding sample covariance matrix is [1.5, 0.16666667], the same as in the old dataset. when converting two datasets into 1-D form, points in their own graph are lies on the same places in Fig.3 they represent same information after principal component analysis. These two datasets represent different information on the original data space, but show the same data characteristics in one dimensional space after PCA process.



(a) Old dataset (b) New dataset

Figure 3: Points in 2 datasets