M12-L2-P1

December 4, 2023

1 M12-L1 Problem 2

Sometimes the dimensionality is greater than the number of samples. For example,in this problem X has 19 features, but there are only 4 data points. You will need to use the alternate PCA formulation in this case. Follow the steps in the cells below to implement this method.

[ ]:
```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.metrics import mean_squared_error

X = np.array([  [-2,  1,  2, -3,  4,  1,  0,  3,  0,  2,  1,  1,  2,  3, -2,␣
 ↪-3, 2,  1,  0],
                [ 1,  2, -4,  2, -4,  2,  5,  2,  2,  1, -3,  0,  0,  1, -2, ␣
 ↪1, 1, -3, -2],
                [ 1, -3,  2,  1,  0, -3, -5, -1,  3,  3, -2, -3, -2, -1,  1, ␣
 ↪0, 5,  4,  2],
                [ 3, -1,  0,  2,  2, -5, -4, -1,  2, -1,  3,  4,  4,  2,  1, ␣
 ↪2, -2,  1, -1]])
```

1.1 Computing Principal Components

1.1.1 The A matrix

First, you should compute the `A` matrix, where A is $(X-\mu)'$. (Note the transpose)

Print this matrix below. It should have size 19×4.

[ ]:
```
A = (X - np.mean(X,axis=0)).T
print("A = \n", A)
```

```
A =
 [[-2.75  0.25  0.25  2.25]
 [ 1.25  2.25 -2.75 -0.75]
 [ 2.   -4.    2.    0.  ]
 [-3.5   1.5   0.5   1.5 ]
 [ 3.5  -4.5  -0.5   1.5 ]
 [ 2.25  3.25 -1.75 -3.75]
 [ 1.    6.   -4.   -3.  ]
 [ 2.25  1.25 -1.75 -1.75]
 [-1.75  0.25  1.25  0.25]
```

```
 [ 0.75 -0.25  1.75 -2.25]
 [ 1.25 -2.75 -1.75  3.25]
 [ 0.5  -0.5  -3.5   3.5 ]
 [ 1.   -1.   -3.    3.  ]
 [ 1.75 -0.25 -2.25  0.75]
 [-1.5  -1.5   1.5   1.5 ]
 [-3.    1.    0.    2.  ]
 [ 0.5  -0.5   3.5  -3.5 ]
 [ 0.25 -3.75  3.25  0.25]
 [ 0.25 -1.75  2.25 -0.75]]
```

1.1.2 "Small" covariance matrix

By transposing $X - \mu$ to get A, now we can compute a smaller covariance matrix with $A'A$. Compute this matrix, `C`, below and print the result.

[ ]:
```
C = A.T@A
print("C = \n", C)
```

```
C =
 [[ 69.875 -18.875 -26.375 -24.625]
 [-18.875 121.375 -53.125 -49.375]
 [-26.375 -53.125  98.375 -18.875]
 [-24.625 -49.375 -18.875  92.875]]
```

1.1.3 Finding nonzero eigenvectors

Next, find the useful (nonzero) eigenvectors of `C`.

For validation purposes, there should be 3 useful eigenvectors, and the first one is `[-0.06628148 -0.79038331 0.47285044 0.38381435]`.

Keep these eigenvectors in a 4×3 array `e`.

[ ]:
```
w,v = np.linalg.eig(C)
w,v = np.real(w), np.real(v)
idx = np.argsort(-w)
w,v = w[idx], v[:,idx]
e = v[:,0:3]
print(e)
```

```
[[-0.06628148  0.04124587 -0.86249959]
 [-0.79038331 -0.06822502  0.34733208]
 [ 0.47285044 -0.69123739  0.22046165]
 [ 0.38381435  0.71821654  0.29470586]]
```

1.1.4 Calculating "eigenfaces"

Now, we have all we need to compute `U`, the matrix of eigenfaces.

$\mathbf{U_i} = \mathbf{Ae_i}$

$(19 \times 3) = (19 \times 4)(4 \times 3)$

Compute and print U. Be sure to normalize your eigenvectors `e` before using the above equation.

[ ]:
```
U = A @ e
U /= np.linalg.norm(U, axis=0)
print("Eigenfaces, U:\n",U)
```

```
Eigenfaces, U:
 [[ 0.07294372  0.12277459  0.33008441]
 [-0.26034151  0.11787331 -0.11677714]
 [ 0.29998485 -0.09606164 -0.27776956]
 [-0.01067529  0.04536213  0.42516696]
 [ 0.27653993  0.17530224 -0.44157072]
 [-0.37621372 -0.15082188 -0.23925816]
 [-0.59257956  0.02265222 -0.05657115]
 [-0.19897063 -0.0037123  -0.250194  ]
 [ 0.04569305 -0.07236581  0.20213547]
 [ 0.0084373  -0.25979087 -0.10504274]
 [ 0.18948616  0.35382298 -0.1518308 ]
 [ 0.00380575  0.46650428 -0.03585222]
 [ 0.03449119  0.40571147 -0.10256065]
 [-0.05241297  0.20419008 -0.19442141]
 [ 0.19396809  0.00756997  0.16057937]
 [ 0.01329023  0.11639359  0.36617258]
 [ 0.0508452  -0.45626561 -0.08985059]
 [ 0.3456779  -0.16842745 -0.07563409]
 [ 0.16171488 -0.18371276 -0.0569842 ]]
```

1.2 Projecting data into 3D

Now project your data into 3 dimensions with the formula:

$ = U^T A $

$(3 \times 4) = (3 \times 19)(19 \times 4)$

Call the projected data Ω "W". Print `W.T`

[ ]:
```
W = U.T@A
print('Projected data in 3 dimensions:\n',W.T)
```

```
Projected data in 3 dimensions:
 [[ -0.8782013    0.44099733  -8.3011616 ]
 [-10.47224127  -0.72945617   3.34291139]
 [  6.26506632  -7.39065157   2.12184196]
 [  5.08537624   7.67911041   2.83640825]]
```

1.3 Reconstructing data in 19-D

We can project the transformed data `W` back into the original 19-D space using:

$\Gamma_f = U\Omega + \Psi$
where:
$ _f = $ reconstructed data
$U = $ eigenfaces
$ = $ Reduced data
$ = $ Means

Do this, and compute the MSE between each reconstructed sample and corresponding original points. Report all 4 MSE values.

[ ]:
```
MSE = []
reconstructed = (U@W) + np.mean(X,axis=1)
mse = np.mean((X - reconstructed.T) **2, axis=1)
for i in range(4):
    print(f"MSE for sample {i}: {mse[i]}")
```

```
MSE for sample 0: 0.7468836565096959
MSE for sample 1: 0.716412742382272
MSE for sample 2: 0.7164127423822705
MSE for sample 3: 0.691481994459834
```

1.4 2-D Reconstruction

What if we had only used the first 2 eigenvectors to compute the eigenfaces? Below, redo the earlier calculations, but use only two eigenfaces. Compute the 4 MSE values that you would get in this case.

(You should get an MSE of 3.626 for the first sample.)

[ ]:
```
MSE2 = []
U2 = A @ e[:,:2]
U2 /= np.linalg.norm(U2, axis=0)
W2 = U2.T@A
reconstructed = (U2@W2) + np.mean(X,axis=1)
print("Using only 2 eigenvectors:")
mse = np.mean((X - reconstructed.T) ** 2, axis=1)
for i in range(4):
    print(f"MSE for sample {i}: {mse[i]}")
```

```
Using only 2 eigenvectors:
MSE for sample 0: 3.882784888090766
MSE for sample 1: 1.3575824909878655
MSE for sample 2: 0.9870175957576055
MSE for sample 3: 1.260330232133847
```