PCA_starter

March 7, 2024

```python
[1]: # Enable interactive rotation of graphs
     %matplotlib widget

     import numpy as np
     from scipy.io import loadmat
     import matplotlib.pyplot as plt
     from mpl_toolkits.mplot3d import Axes3D

     # Load data for activity
     X = loadmat('PCA_Activity.mat')['X']
     rows, cols = np.array(X.shape)
     x, y, z = X

     print('Rows of X = ',rows)
     print('Cols of X = ',cols)
```
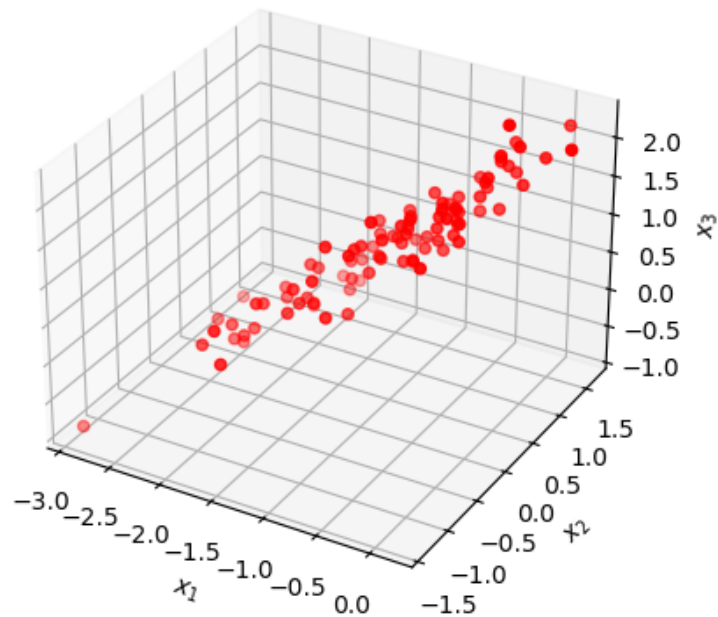
```
Rows of X =  3
Cols of X =  100
```

```python
[2]: fig = plt.figure()
     ax = fig.add_subplot(111, projection='3d')

     ax.scatter(x, y, z, c='r', marker='o')

     ax.set_xlabel('$x_1$')
     ax.set_ylabel('$x_2$')
     ax.set_zlabel('$x_3$')

     plt.show()
```

0.0.1 a) Run the code to display the data in Figure the first figure. Use the rotate tool to inspect the scatter plot from different angles. Does the data appear to lie very close to a one-dimensional subspace? Does the data appear to be zero mean?

Yes, the data appears to lie very close to 1-dim subspace. No, it doesn't appear to be zero mean, because the center of the cluster seems to be around (-1.5, 1, 0.5).

```
[3]: # Subtract mean
     X_m = X #- np.mean(X, 1).reshape((3,1))
     X_m = X - np.mean(X, 1).reshape((3,1))
     x_m, y_m, z_m = X_m
```

0.0.2 c) Now comment out (insert %) the line of code that subtracts the mean of the data. Does the dominant feature identified by SVD continue to be a good fit to the data? Comment on the importance of removing the mean before performing PCA.

No, it does not continue to be a good fit to the data anymore. Thus, we know it is important to remove the mean before performing PCA.
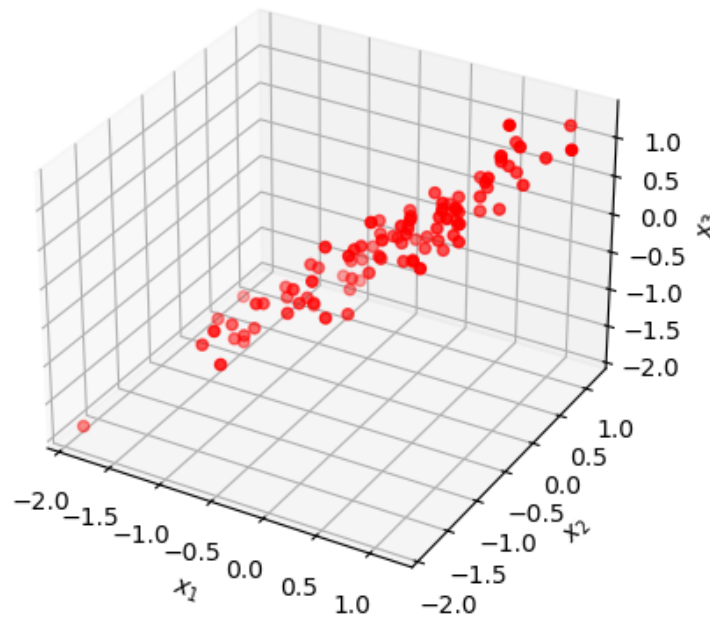
```
[4]: # display zero mean scatter plot

     fig = plt.figure()
     ax = fig.add_subplot(111, projection='3d')

     ax.scatter(x_m, y_m, z_m, c='r', marker='o')

     ax.set_xlabel('$x_1$')
     ax.set_ylabel('$x_2$')
     ax.set_zlabel('$x_3$')

     plt.show()
```


0.0.3 b) Figure 2 depicts the centered data and the one-dimensional subspace that contains the dominant feature you identified using the SVD. Use the rotate tool to inspect the data and one-dimensional subspace from different angles. Is a one-dimensional subspace a reasonable fit to the data? Comment on the error.

1-dim subspace can partly reasonable fit to the data, but it cannot capture some features which data spanning on the direction orthoganal to the subspace contains.

3

```
[5]:  # Use SVD to find first principal component

      U,s,VT = np.linalg.svd(X_m,full_matrices=False)

      # complete the next line of code to assign the first principal component to a
      a =  U[:,0]

      print(a)
```

```
[-0.58277194 -0.57701087 -0.57221964]
```

```
[6]:  # display zero mean scatter plot and first principal component

      fig = plt.figure()
      ax = fig.add_subplot(111, projection='3d')

      ax.scatter(x_m, y_m, z_m, c='r', marker='o', label='Data')

      ax.scatter(a[0],a[1],a[2], c='c', marker='s')

      ax.set_xlabel('$x_1$')
      ax.set_ylabel('$x_2$')
      ax.set_zlabel('$x_3$')

      ax.plot([0,a[0]],[0,a[1]],[0,a[2]], c='b',label='Principal Component')

      ax.legend()
      plt.show()
```
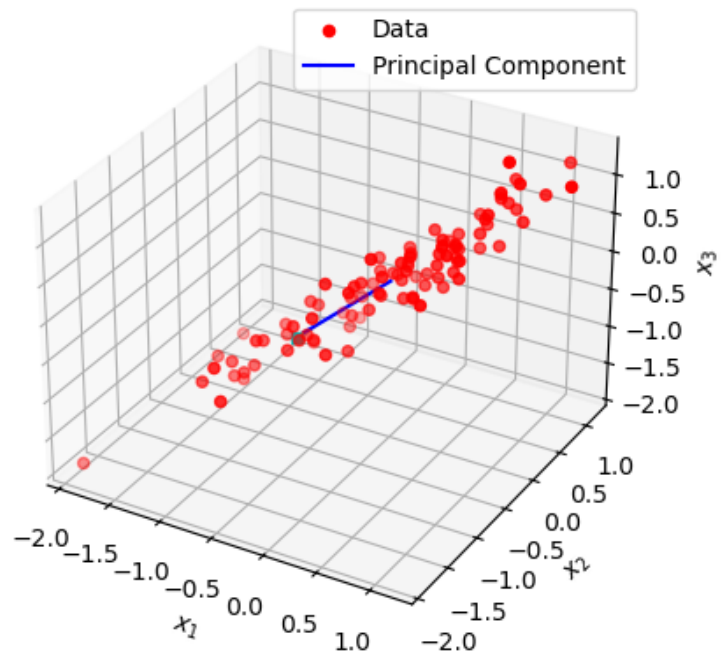
Activity 2

Item 1.

1. X, $n\times p$ (rank r), $= U\Sigma V^T$, $n\times r$, $r\times r$, $r\times p$, $n>p>r$

a) $Z = X^T$

$\to Z = (U\Sigma V^T)^T = V\Sigma^T U^T = V\Sigma U^T$

b) rank 1   rows of Z = col of X

subspace: u_1

2. X

$n\times p$, $p<n$

a) rank$(X) \neq p$, not full rank
$(<)$

b) $\min_w \| Xw - y\|_2^2 + \lambda \|w\|_2^2$

$\to \min_w \left\| \begin{bmatrix} Xw-y \\ \lambda^{1/2} w \end{bmatrix} \right\|_2^2 \to \min_w \left\| \begin{bmatrix} X \\ \lambda^{\frac{1}{2}} I \end{bmatrix} w - \begin{bmatrix} y \\ 0 \end{bmatrix} \right\|_2^2$

$\tilde{X}$, $\tilde{y}$

c) $\lambda > 0$
rank$(\hat{X}) = $ rank$([\hat{X}\, \hat{y}]) = p$ full rank

3. a) $X \in R^{n\times p}$, $p \leq n$

① $X = U\Sigma V^T = \sum_{i=1}^{p} \sigma_i u_i v_i^T$

$\left((\sigma_i u_i v_i^T)^T (\sigma_i u_i v_i^T) + \lambda I \right)^{-1} (\sigma_i u_i v_i^T)$

$= \left((v_i u_i^T \sigma_i)(\sigma_i u_i v_i^T) \right)$

$(\sigma_i^2 + \lambda I)^{-1} (\sigma_i v_i u_i^T \, \sigma_i)$

$= (\sigma_i^2 + \lambda)^{-1} (\sigma_i v_i v_i u_i^T)$

$= \dfrac{\sigma_i}{\sigma_i^2 + \lambda} v_i u_i^T$

② $(X^T X + \lambda I)^{-1} X^T$

$= (V\Sigma^2 V^T + \lambda I)(U\Sigma V^T)^T$

$= (V(\Sigma^2 + \lambda I) V^T)^{-1}$

$= (V^{-1}\Sigma^{-2} V^{-1} + V^{-1}\tfrac{1}{\lambda} I V^{-1})$

$(V\Sigma^{-2}V^T + V\tfrac{1}{\lambda} \Sigma V^T)$

$\to V(\Sigma^{-2}+\tfrac{1}{\lambda}) X^T \Sigma U^T$

$= \Sigma^{2\times\lambda} V U^T$

$A^T = A^T$

b) SVD $X = U\Sigma V^T$

$X^T X = V\Sigma U^T U \Sigma V^T = V\Sigma^2 V^T$

$\underbrace{\quad}_{Q}$

$X^+ = \lim_{\lambda\to 0} (X^T X + \lambda I)^{-1} X^T$ if X invertible

$\to (X^T X)^{-1} X^T$ $\lambda = 0$

c) X is square and invertible

$n\times n$ $X^+ = \boxed{(X^T X)^{-1} X^T}$

$= \underbrace{X^{-1} X^{-T} X^T}_{I}$

$= X^{-1}$

d) X rank $r < p$

$\sigma_1 \geq \sigma_2 \geq \cdots \sigma_r \geq \sigma_{r+1} = \cdots \sigma_p = 0$

So $\sum_{i=1}^{r} \dfrac{\sigma_i}{\sigma_i^2 + \lambda} v_i u_i^T$ $r<p$

e) $X^+ = X^{-1}$

$= (U\Sigma V^T)^{-1}$

$= V^{-1}\Sigma^{-1}U^{-1}$

$= V\Sigma^{-1}U^T$

$\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1} & & & & 0 \\ & \frac{1}{\sigma_2} & \frac{1}{\sigma_3} \cdots & & \frac{1}{\sigma_r} \\ & & & \ddots & \frac{1}{\sigma_p} \\ 0 & & & & =0 \end{bmatrix}$