Stat601(Section001): Homework #1

Due on Jan. 19, 2021 at 11:59pm

Instructor:Ziwei Zhu

Tiejin Chen **tiejin@umich.edu**

Problem 1

We get the estimation of Θ is:

$$
\begin{bmatrix}
-3.e-05 & 2.1e-05 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
1.3e-05 & -9.e-06 & 0.0 & 2.5e-08 & 0.0 & 0.0 & 1.2e-07 & -1.e-08 & 0.0 & 1.6e-08 & 0.0 \\
0.0 & 0.0 & 2.34e-4 & 1.22e-4 & 0.0 & 0.0 & 0.0 & 1.5e-06 & 0.0 & -4.e-06 & 0.0 \\
0.0 & 2.6e-07 & 3.3e-05 & -2.e-05 & 0.0 & 0.0 & 2.4e-06 & 5.2e-08 & 0.0 & -5.e-07 & 1.3e-06 \\
0.0 & 0.0 & 0.0 & 0.0 & 3.3e-05 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 3.3e-05 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & -2.e-06 & 0.0 & -5.e-06 & 0.0 & 0.0 & 3.9e-05 & 0.0 & 0.0 & -2.e-06 & 0.0 \\
0.0 & 3.6e-05 & 1.27e-4 & -1.e-05 & 0.0 & 0.0 & 0.0 & 2.08e-4 & 0.0 & 2.8e-05 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 4.4e-05 & -8.e-06 & 0.0 \\
0.0 & 5.6e-08 & 3.7e-07 & -1.e-07 & 0.0 & 0.0 & 4.5e-07 & -3.e-08 & 2.2e-05 & -5.e-06 & 1.7e-06 \\
0.0 & 0.0 & 0.0 & 1.21e-3 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 5.25e-3 & 1.55e-2
\end{bmatrix}
$$

Problem 2

We use Lagrange multiplier to solve this problem. And we can get:

$$L(X_i, \mu, U, \Lambda) = \sum_{i=1}^{N} \|Y_i - \mu - UX_i\|^2 + \lambda(U^T U - I_p)$$

And we get partial derivatives of λ and X_i and let it to 0:

$$\frac{\partial L}{\partial \lambda} = U^T U - I_p = 0$$

$$\frac{\partial L}{\partial X_i} = (Y_i - \mu - UX_i)U^T = 0 \to Y_i - \mu = UX_i \to U^T(Y_i - \mu) = X_i$$

Then we take partial derivatives of μ to get:

$$\frac{\partial L}{\partial \mu} = \sum_{i=1}^{N} -2(Y_i - \mu - Ux_i) = 0 \to \mu = \bar{Y} - U\bar{X}$$

Here $X \sim N_p(0, I_p)$ is a latent variable. And $\bar{X}$ should equal to the mean of X. Hence we can get:

$$\bar{X} = 0$$

$$\mu = \bar{Y}$$

$$X_i = U^T(Y_i - \bar{Y})$$

We take partial derivatives of U to get:

$$\frac{\partial L}{\partial U} = -\sum_{i=1}^{N}(Y_i - \bar{Y} - UX_i)X_i^T + 2\lambda U = 0 \to (\lambda + \sum_{i=1}^{N} X_i X_i^T)U = \sum_{i=1}^{N}(Y_i - \bar{Y})X_i^T$$

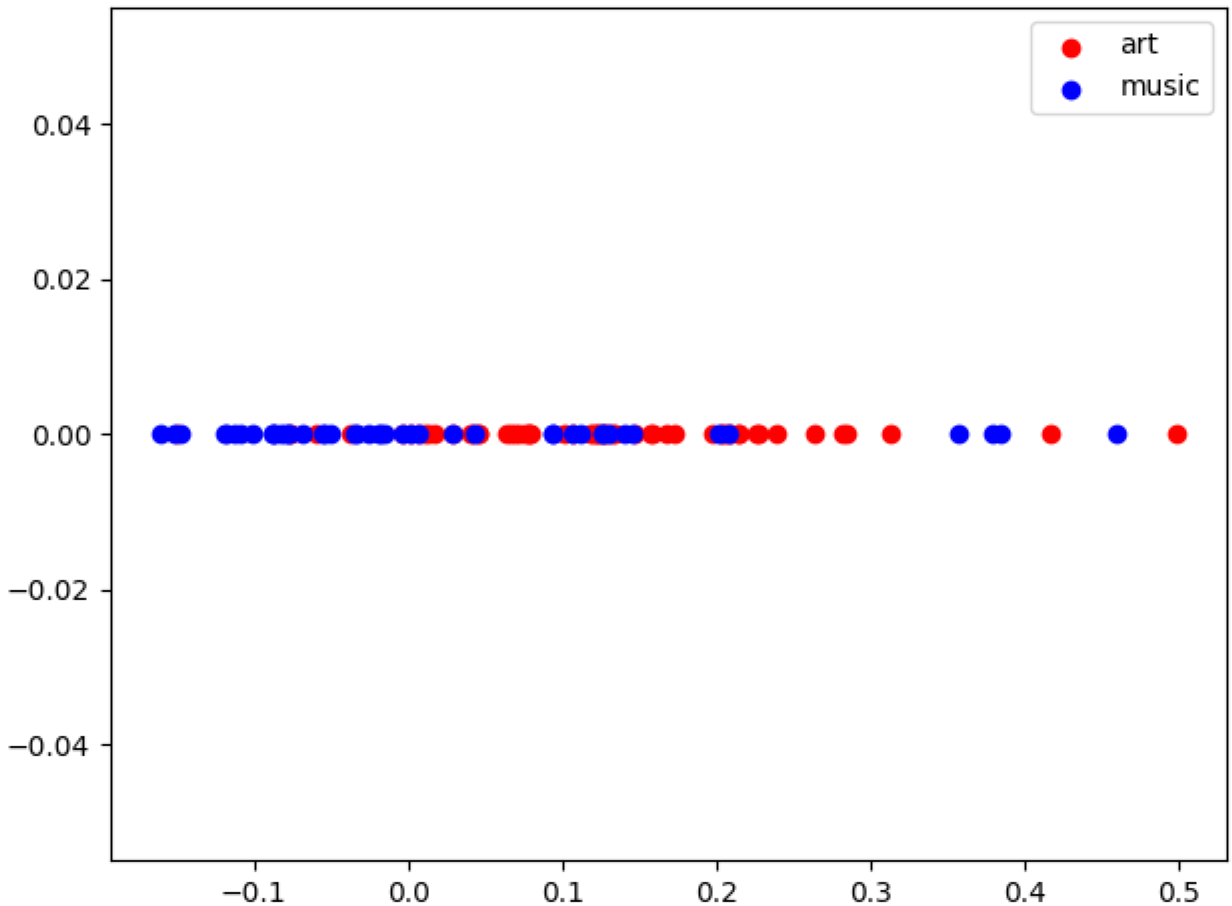
Since $\lambda + \sum_{i=1}^{N} X_i X_i^T$ is a scalar, we leave it and plug $X_i = U^T(Y_i - \bar{Y})$ in the RHS to get:

$$(\lambda + \sum_{i=1}^{N} X_i X_i^T)U = \sum_{i=1}^{N}(Y_i - \bar{Y})(Y_i - \bar{Y})^T U = (N-1)SU \to \frac{(\lambda + \sum_{i=1}^{N} X_i X_i^T)}{N-1}U = SU \to \theta U = SU$$

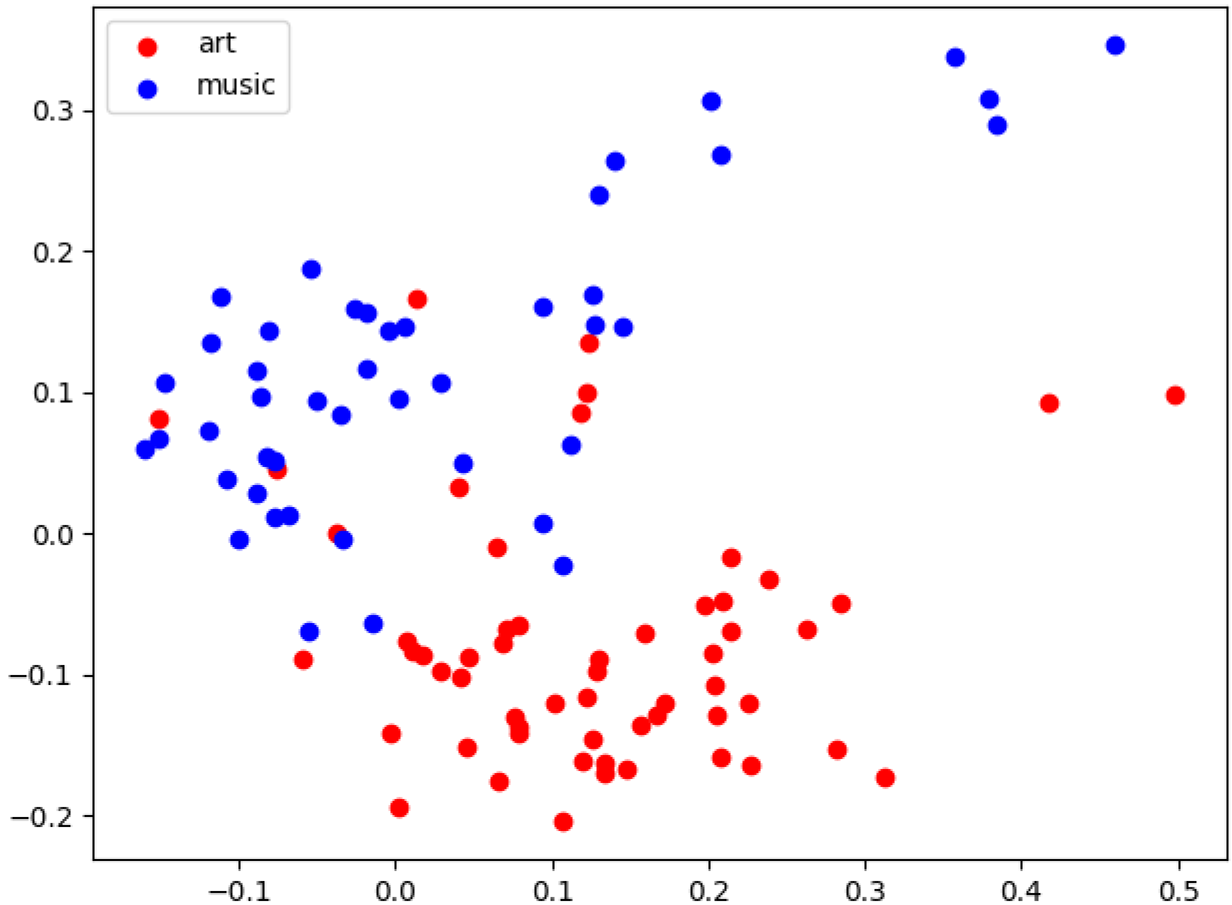
Hence U contains top p eigenvalues of sample variance matrix of Y, Which is the end of our prove. Actually it is not hard to see that $\lambda = 0$ since $\sum_{i=1}^{N}(Y_i - \bar{Y} - UX_i)X_i^T$ is zero. And $\sum_{i=1}^{N} X_i X_i^T = (N-1)U^T SU = (N-1)\phi$ is N-1 times the eigenvalues of S.

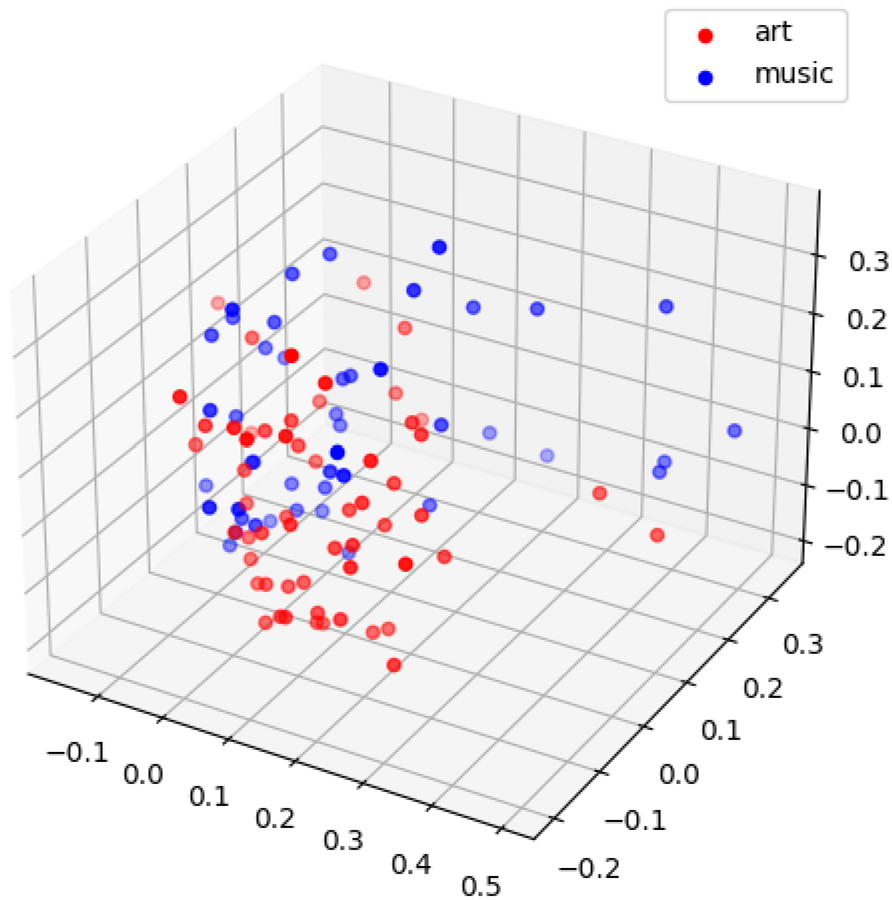
Problem 3

(a)
Firstly, we started with Schmidt orthogonalization and next QR decomposition and then getting eigenvalues and eigenvectors. we use some more basic functions from numpy such as np.linalg.inv() or np.dot(). Then we found my code runs very slow for a 4000 by 4000 matrix. Hence we start to use np.eig() to get the eigenvalue and eigenvector results. See detail in our codes.
(b)
We get the plots:



Problem 3 continued on next page. . .


art
music
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
0.5



2-dim might be enough for us to visually separate the two classes of articles reasonably well, though it might be some mistakes.

(c)

1. W_1
   - 20 words with maximum positive weights:['she', 'her', 'ms', 'i', 'said', 'mother', 'cooper', 'my', 'painting', 'process', 'paintings', 'im', 'he', 'mrs', 'me', 'gagosian', 'was', 'picasso', 'image', 'sculpture']
   - 20 words with maximum negtive weights:['music', 'trio', 'theater', 'orchestra', 'composers', 'opera', 'theaters', 'm', 'festival', 'east', 'program', 'y', 'jersey', 'players', 'committee', 'sunday', 'june', 'concert', 'symphony', 'organ']
2. W_2
   - 20 words with maximum positive weights:['her', 'she', 'theater', 'opera', 'ms', 'i', 'hour', 'production', 'sang', 'festival', 'music', 'musical', 'songs', 'vocal', 'orchestra', 'la', 'singing', 'matinee', 'performance', 'band']
   - 20 words with maximum negtive weights:['art', 'museum', 'images', 'artists', 'donations', 'museums', 'painting', 'tax', 'paintings', 'sculpture', 'gallery', 'sculptures', 'painted', 'white', 'patterns', 'artist', 'nature', 'service', 'decorative', 'feet']

3. W_3

- 20 words with maximum positive weights:['said', 'theater', 'will', 'museum', 'museums', 'donations', 'art', 'tax', 'national', 'directors', 'law', 'director', 'm', 'theaters', 'jersey', 'yesterday', 'million', 'new', 'dealers', 'festival']
- 20 words with maximum negtive weights:['trio', 'band', 'her', 'songs', 'music', 'orchestra', 'bands', 'sound', 'guitarist', 'guitar', 'ms', 'sounds', 'patterns', 'review', 'sometimes', 'seems', 'vocal', 'la', 'red', 'spatial']

For me W_1 capture the information about art, W_2 captures the information about music, W_3 captures the information about time and place.

Problem 4

(a)
let $\hat{\Lambda} = \Lambda C, \hat{X} = C^{-1}X$. Thus We have:

$$Y = \hat{\Lambda}\hat{X} + W + \mu = \Lambda X + W + \mu$$

This two model are equivalent and we can not distinguish which Λ is right. However, if we let X and C to be orthogonal. Then we have:

$$CC^{-1}XX^TC^{-T}C^T = I$$

Thus $CC^{-1}X$ is also an orthogonal matrix. Then the model will be:

$$Y = \Lambda X + W + \mu, XX^T = I$$

It will get a distinguishable Λ. Thus we prove Λ can only be identified up to an orthornormal transformation. For measure to the evaluate the distance of two span linear space, the space will differ only when the rank of Λ and Λ^* are different since they have same shape. Hence we will use $D(span(\Lambda), span(\Lambda^*)) = |rank(\Lambda) - rank(\Lambda^*)|$
(b)
We know the joint pdf of (X, Y) is

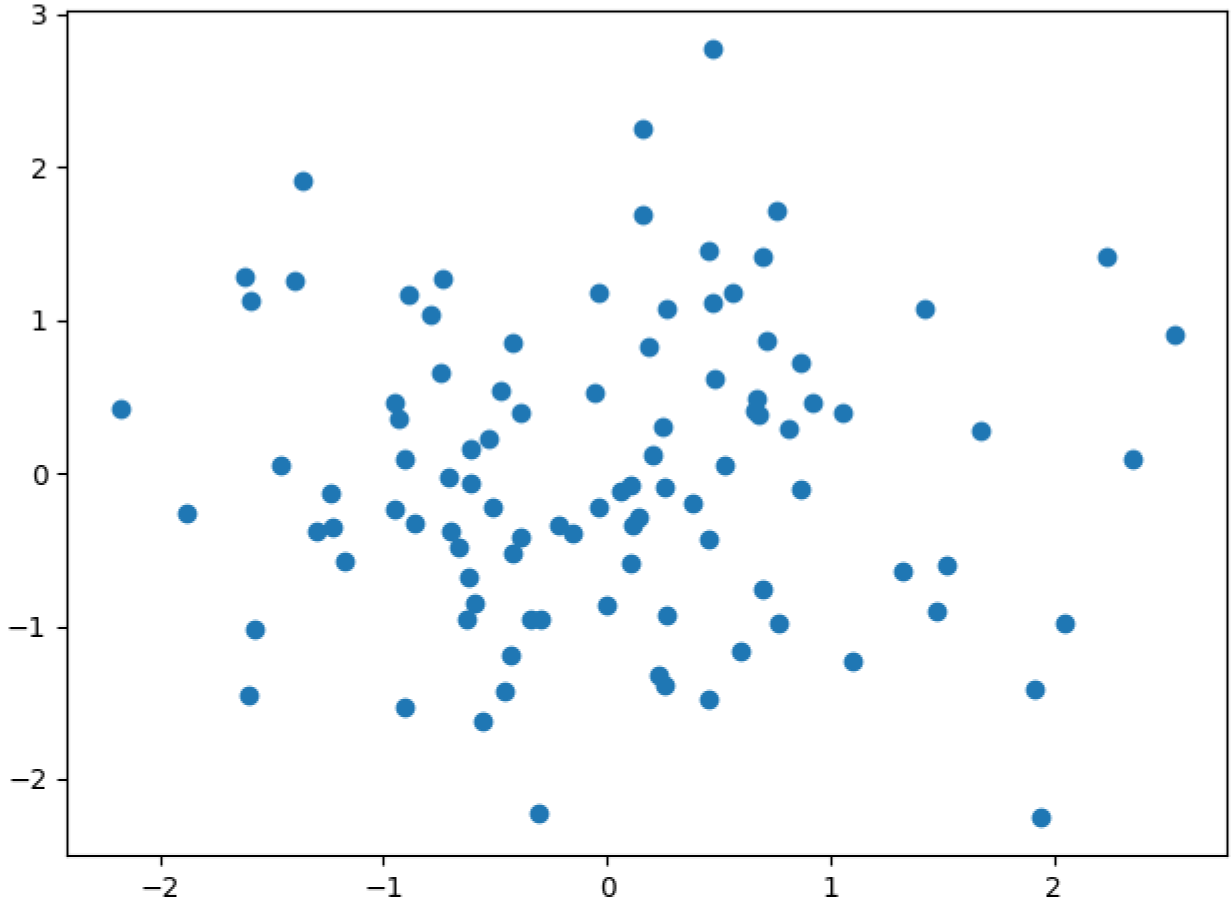
$$(X, Y) \sim N\begin{pmatrix} 0 \\ \mu^* \end{pmatrix}\begin{pmatrix} I & \Lambda^{*T} \\ \Lambda^* & \Psi + \Lambda^*\Lambda^{*T} \end{pmatrix}$$

Thus we can get the mean of $X|Y$ is $\Lambda^{*T}(\Psi + \Lambda^*\Lambda^{*T})^{-1}(Y - \mu^*)$. We can get the MLE of $X|Y$ is:

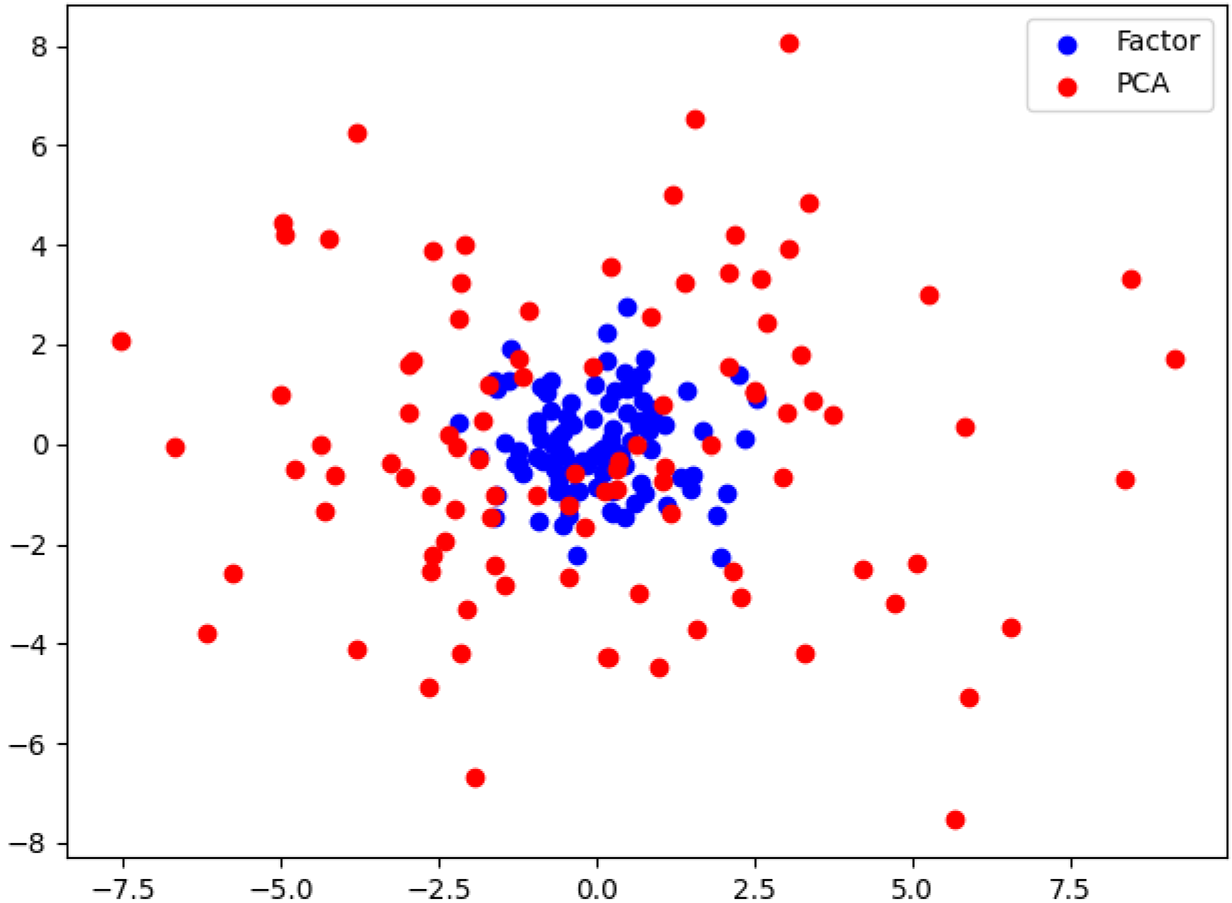
$$\Lambda S^{-1}(Y - \bar{Y})$$

And $(Y - \bar{Y})$ can been seen as the coordinate in the subspace. Thus ΛS^{-1} defines the subspace of X where Λ is the estimated Λ^* and S is the corvariance matrix of Y.
We will use FactorAnlysis function from sklearn to get the result.

Problem 4 continued on next page...



(c)
We will use FactorAnlysis and PCA function from sklearn and we get:



We can see that the result of factor analysis are much more compact than PCA result. But they have similar distribution of data point.

Problem 5

(a)
When we set time of procedure is 500, the number of PC components is 2
(b)
We set time of procedure is 500,And the simulation times is 100. permutation test estimate number of components correctly for every time.
(c)
We set the simulation times to 100. And it never estimate number of components correctly. The estimated number of components always greater than 2.
(d)
in this part, we use FactorAnlysis from sklearn. And see the detail function likelihood_ra_test in the codes.
(e)
We also set the simulation times to 100. And we get the correct ratio of likelihood ratio test is 0.9.