Compute PCA by hand for the following data matrix

$$X = \begin{bmatrix} -3 & -2 \\ -2 & -1 \\ -1 & -1 \\ 1 & 1 \\ 2 & 1 \\ 3 & -1 \end{bmatrix}$$

1). Find the mean vector μ^T (size = 1x2)

0 -0.5

2). Form a 6x2 mean matrix: $M = \begin{bmatrix} \mu^T \\ \mu^T \\ \mu^T \\ \mu^T \\ \mu^T \\ \mu^T \end{bmatrix}$

0 -0.5000
0 -0.5000
0 -0.5000
0 -0.5000
0 -0.5000
0 -0.5000

3). Create a 6x2 zero-mean data matrix: $XM = X - M$

-3.0000 -1.5000
-2.0000 -0.5000
-1.0000 -0.5000
1.0000 1.5000
2.0000 1.5000
3.0000 -0.5000

4). Compute the 2x2 unbiased covariance matrix: $\Sigma = \frac{1}{N-1} XM^T . XM$

- N = the number of observations

5.6000 1.8000
1.8000 1.5000

5). Find the eigenvalues λ_1, λ_2 and the eigenvectors v_1, v_2 of the covariance matrix Σ

- $\lambda_1 > \lambda_2$ and v_1, v_2 are 2x1vectors

Eigenvalues	Eigenvectors
$\lambda_1 = 6.2781$	$v_1 =$ -0.9358 -0.3525
$\lambda_2 = 0.8219$	$v_2 =$ 0.3525 -0.9358

6). Form a 2x2 eigen matrix consisting of the two eigenvectors: $\Phi = [v_1 \quad v_2]$

-0.9358 0.3525
-0.3525 -0.9358

7). Compute the principal components by projecting XM onto Φ: $Y = XM.\Phi$

- Y is a 6x2 matrix. The first column is PC1 and the second column is PC2

3.3362 0.3461
2.0479 -0.2372
1.1121 0.1154
-1.4646 -1.0512
-2.4004 -0.6986
-2.6311 1.5255

8). Compute the reconstructed data matrix: $\hat{X} = \bar{Y}.\Phi^T + M$

(a). Use only PC1 → $\bar{Y}$ is a 6x1 vector which is the first column of Y and Φ is the 1st eigenvector $\Phi = v_1$

-3.1220	-1.6761
-1.9164	-1.2219
-1.0407	-0.8920
1.3706	0.0163
2.2463	0.3462
2.4622	0.4276

(b). Compute the reconstruction error

$$\text{Error} = \frac{\|X - \hat{X}\|_F}{\|X\|_F} = \frac{\left\| \begin{matrix} 0.1220 & -0.3239 \\ -0.0836 & 0.2219 \\ 0.0407 & -0.1080 \\ -0.3706 & 0.9837 \\ -0.2463 & 0.6538 \\ 0.5378 & -1.4276 \end{matrix} \right\|_F}{\left\| \begin{matrix} -3 & -2 \\ -2 & -1 \\ -1 & -1 \\ 1 & 1 \\ 2 & 1 \\ 3 & -1 \end{matrix} \right\|_F} = \frac{2.0272}{6.0828} = 0.3333$$

(c) Use both PC1 and PC2. Hence, $\bar{Y} = Y$ and $\Phi = [v_1 \quad v_2]$

-3.0000	-2.0000
-2.0000	-1.0000
-1.0000	-1.0000
1.0000	1.0000
2.0000	1.0000
3.0000	-1.0000

(d). Compute the reconstruction error

$$\text{Error} = \frac{\|X - \hat{X}\|_F}{\|X\|_F} = \frac{0}{6.0828} = 0$$