PCA example:

Let we have a dataset with four observations and two features:

X1 X2

---------

1 2

2 3

3 4

4 5

Here's how we'll perform PCA step by step:

**Step 1: Data Standardization**

First, we standardize the dataset by subtracting the mean and dividing by the standard deviation for each feature:

X1\_std = (1 - 2.5) / 1.29 ≈ -0.39

= (2 - 2.5) / 1.29 ≈ -0.39

= (3 - 2.5) / 1.29 ≈ 0.39

= (4 - 2.5) / 1.29 ≈ 1.56

X2\_std = (2 - 3.5) / 1.29 ≈ -0.78

= (3 - 3.5) / 1.29 ≈ -0.39

= (4 - 3.5) / 1.29 ≈ 0.39

= (5 - 3.5) / 1.29 ≈ 1.17

So, the standardized dataset looks like this:

X1\_std X2\_std

----------------

-0.39 -0.78

-0.39 -0.39

0.39 0.39

1.56 1.17

**Step 2: Covariance Matrix Calculation**

Next, we compute the covariance matrix for the standardized dataset:

cov(X1\_std, X1\_std) cov(X1\_std, X2\_std)

cov(X2\_std, X1\_std) cov(X2\_std, X2\_std)

1.00 0.83

0.83 1.00

**Step 3: Eigen Decomposition**

We perform eigen decomposition on the covariance matrix to obtain the eigenvalues and eigenvectors.

Eigenvalues (λ):

λ1 ≈ 1.83

λ2 ≈ 0.17

Eigenvectors (v):

v1 ≈ [0.71, 0.71]

v2 ≈ [-0.71, 0.71]

**Step 4: Dimensionality Reduction**

We select the principal components based on the eigenvalues. Since λ1 > λ2, we choose the first eigenvector (v1) as the principal component.

**Step 5: Projection**

Finally, we project the standardized data onto the first principal component:

X\_pca = X\_std.dot(v1)

= [0.39, 0.39, -0.39, 1.56] \* [0.71, 0.71]

≈ [0.55, 0.28, -0.28, -1.11]

So, the transformed data in the reduced-dimensional space is:

X\_pca

-------

0.55

0.28

-0.28

-1.11

This represents the original dataset projected onto the first principal component, capturing the maximum variance in the data.