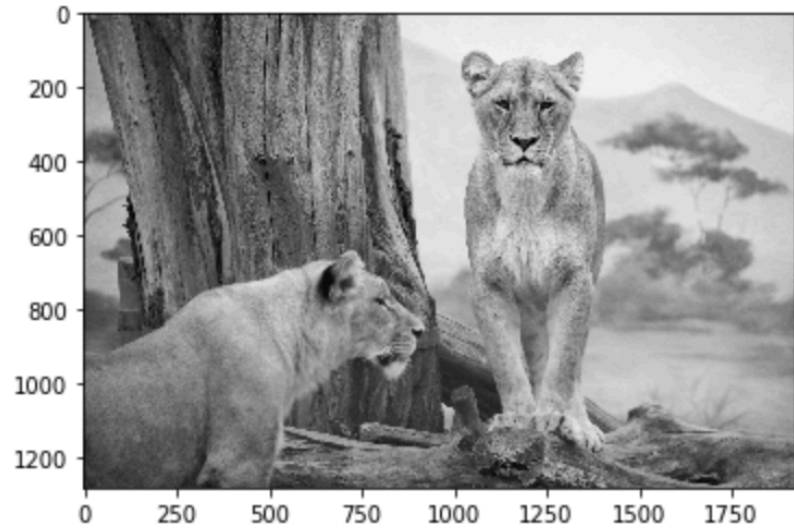


SINGULAR VALUE DECOMPOSITION (SVD)/ PRINCIPAL COMPONENTS ANALYSIS (PCA)

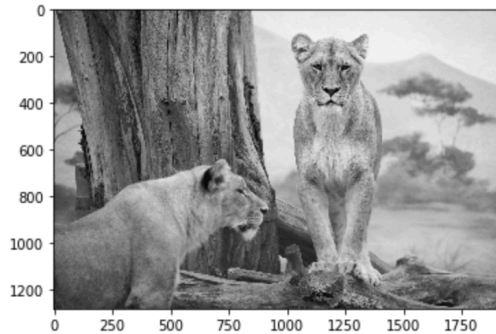
SVD - EXAMPLE



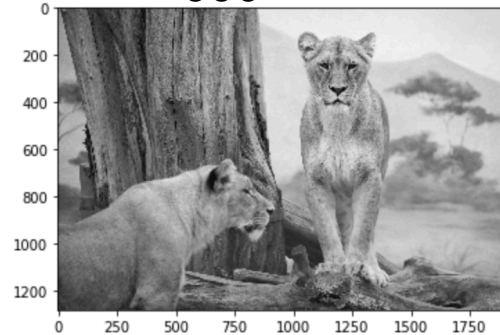
$$U, S, V^T = \text{numpy.linalg.svd}(img)$$

SVD - EXAMPLE

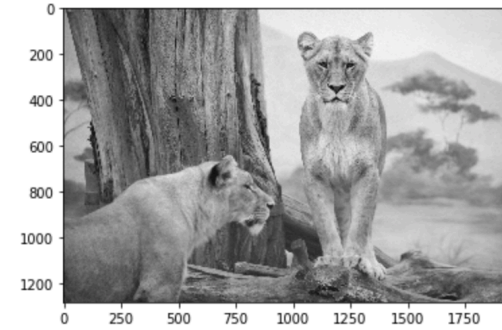
full rank



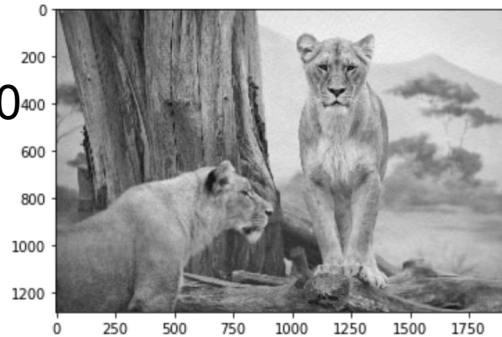
600



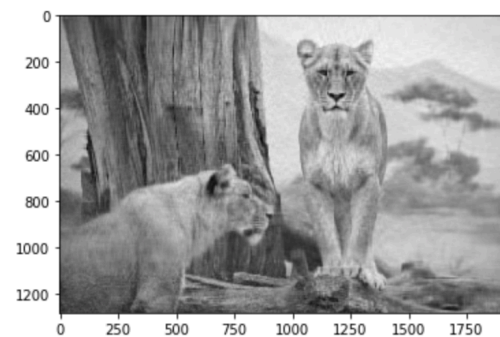
300



100



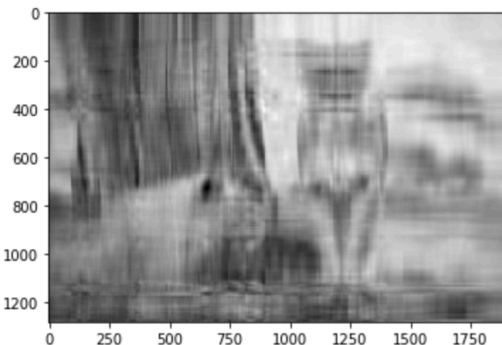
50



20



10



$$U[:, k]S[:, k]V^T[:, k, :]$$

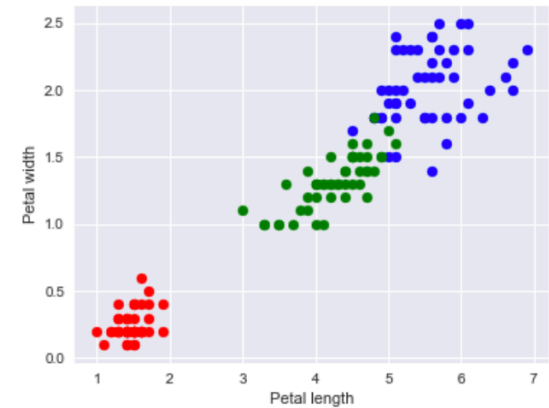
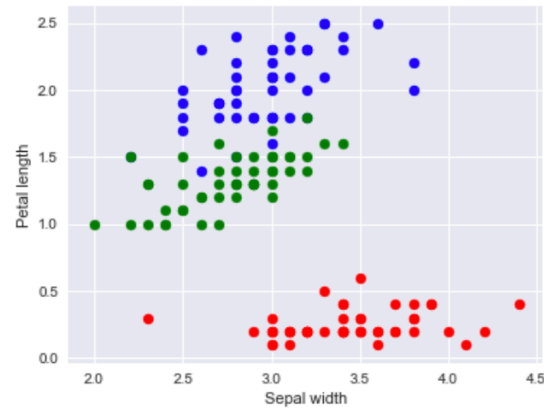
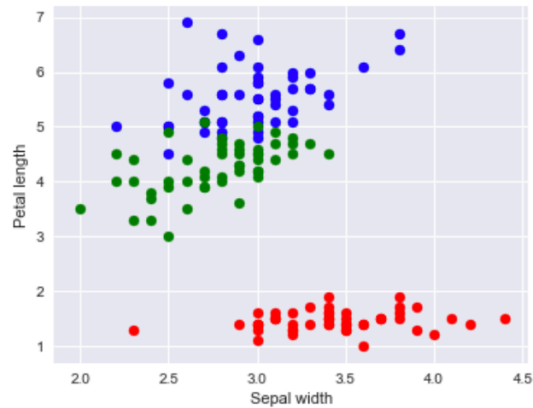
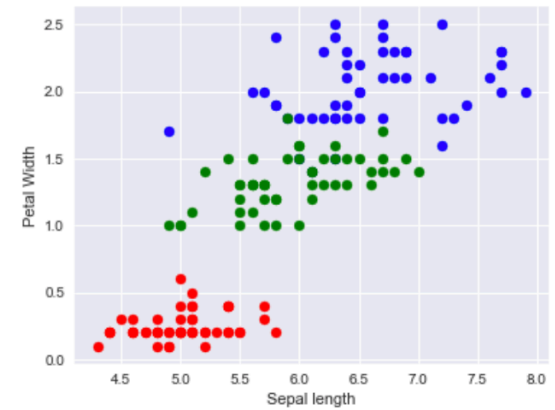
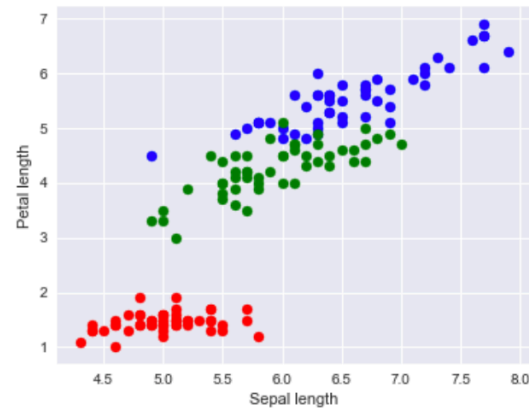
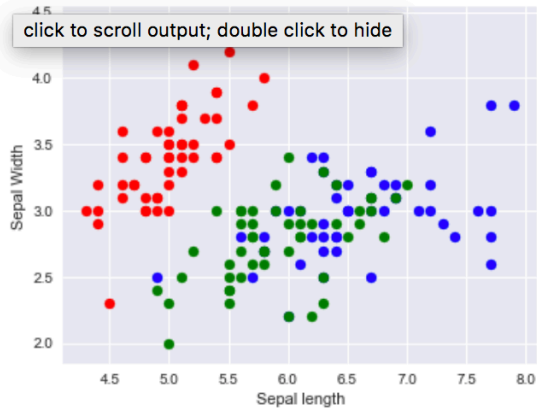
PCA - INTRODUCTION

$$X = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 1 & 5 \\ 3 & 4 & 10 \\ 4 & 3 & 11 \end{bmatrix}$$

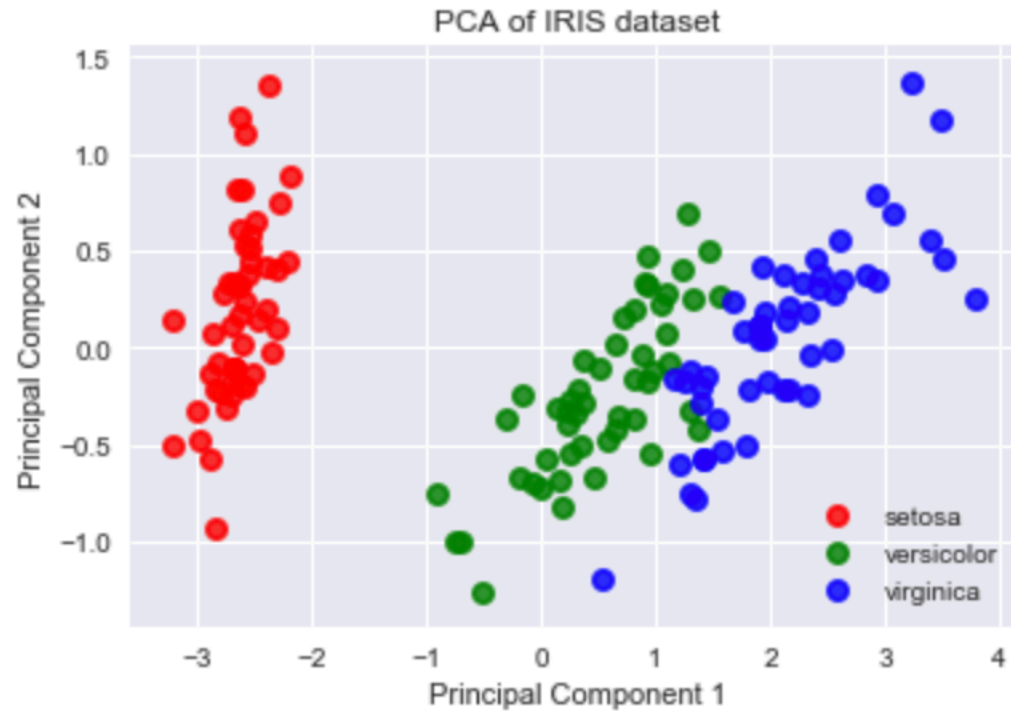
PCA - INTRODUCTION

	sepal_length	sepal_width	petal_length	petal_width	species
8	4.4	2.9	1.4	0.2	setosa
52	6.9	3.1	4.9	1.5	versicolor
35	5.0	3.2	1.2	0.2	setosa
127	6.1	3.0	4.9	1.8	virginica
96	5.7	2.9	4.2	1.3	versicolor

PCA - INTRODUCTION



PCA - INTRODUCTION



PRINCIPAL COMPONENT ANALYSIS

- A technique to find the directions along which the points (set of tuples) in high-dimensional data line up best.
- Treat a set of tuples as a matrix M and find the eigenvectors for MM^T or M^TM .
- The matrix of these eigenvectors can be thought of as a rigid rotation in a high-dimensional space.
- When this transformation is applied to the original data - the axis corresponding to the principal eigenvector is the one along which the points are most “spread out”.

PRINCIPAL COMPONENT ANALYSIS

- When this transformation is applied to the original data - the axis corresponding to the principal eigenvector is the one along which the points are most “spread out”.
- This axis is the one along which variance of the data is maximized.
- Points can best be viewed as lying along this axis with small deviations from this axis.
- Likewise, the axis corresponding the second eigenvector is the axis along which the variance of distances from the first axis is greatest, and so on.

PRINCIPAL COMPONENT ANALYSIS

- **Principal Component Analysis (PCA) is a dimensionality reduction method.**
- **The goal is to embed data in high dimensional space, onto a small number of dimensions.**
- **It most frequent use is in exploratory data analysis and visualization.**
- **It can also be helpful in regression (linear or logistic) where we can transform input variables into a smaller number of predictors for modeling.**

PRINCIPAL COMPONENT ANALYSIS

- Mathematically,

Given: Data set $\{x_1, x_2, \dots, x_n\}$

where, x_i is the vector of p variable values for the i -th observation.

Return:

Matrix $[\phi_1, \phi_2, \dots, \phi_p]$

of linear transformations that retain maximal variance.

- **You can think of the first vector ϕ_1 as a linear transformation that embeds observations into 1 dimension**

$$Z_1 = \phi_{11}X_1 + \phi_{21}X_2 + \dots + \phi_{p1}X_p$$

PRINCIPAL COMPONENT ANALYSIS

- You can think of the first vector ϕ_1 as a linear transformation that embeds observations into 1 dimension

$$Z_1 = \phi_{11}X_1 + \phi_{21}X_2 + \dots + \phi_{p1}X_p$$

where ϕ_1 is selected so that the resulting dataset $\{z_i, \dots, z_n\}$ has maximum variance.

- In order for this to make sense, mathematically, data has to be centered
 - Each X_i has zero mean
 - Transformation vector ϕ_1 has to be normalized, i.e., $\sum_{j=1}^p \phi_{j1}^2 = 1$

PRINCIPAL COMPONENT ANALYSIS

- In order for this to make sense, mathematically, data has to be centered
 - Each X_i has zero mean
 - Transformation vector ϕ_1 has to be normalized, i.e., $\sum_{j=1}^p \phi_{j1}^2 = 1$
- We can find ϕ_1 by solving an optimization problem:

$$\max_{\phi_{11}, \phi_{21}, \dots, \phi_{p1}} \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^p \phi_{j1} x_{ij} \right)^2 \text{ s.t. } \sum_{j=1}^p \phi_{j1}^2 = 1$$

Maximize variance but subject to normalization constraint.

PRINCIPAL COMPONENT ANALYSIS

- We can find ϕ_1 by solving an optimization problem:

$$\max_{\phi_{11}, \phi_{21}, \dots, \phi_{p1}} \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^p \phi_{j1} x_{ij} \right)^2 \text{ s.t. } \sum_{j=1}^p \phi_{j1}^2 = 1$$

Maximize variance but subject to normalization constraint.

- The second transformation, ϕ_2 is obtained similarly with the added constraint that ϕ_2 is orthogonal to ϕ_1
- Taken together $[\phi_1, \phi_2]$ define a pair of linear transformations of the data into 2 dimensional space

$$Z_{n \times 2} = X_{n \times p} [\phi_1, \phi_2]_{p \times 2}$$

PRINCIPAL COMPONENT ANALYSIS

- Taken together $[\phi_1, \phi_2]$ define a pair of linear transformations of the data into 2 dimensional space

$$Z_{n \times 2} = X_{n \times p}[\phi_1, \phi_2]_{p \times 2}$$

- Each of the columns of the Z matrix are called Principal components.
- The units of the PCs are meaningless.
- In practice we may also scale X_j to have unit variance.
- In general if variables X_j are measured in different units(e.g., miles vs. liters vs. dollars), variables should be scaled to have unit variance.

SPECTRAL THEOREM

Using Spectral theorem

$$(X^T X)\phi = \lambda\phi$$

$$XX^T(X\phi) = \lambda(X\phi)$$

Conclusion:

The matrices XX^T and $X^T X$ share the same nonzero eigenvalues

To get an eigenvector of XX^T from $X^T X$ multiply ϕ on the left by X

Very powerful, particularly if number of observations, m , and the number of predictors, n , are drastically different in size.

For PCA:

$$\text{Cov}(X, X) = XX^T$$

EXAMPLE - PCA

$$X = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix}$$

Eigen Values and Eigen Vectors?

EXAMPLE - PCA

$$X = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix}$$

From spectral theorem:

$$(X^T X)\phi = \lambda\phi$$

$$X^T X = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 30 & 28 \\ 28 & 30 \end{bmatrix}$$

EXAMPLE - PCA

$$X = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix}$$

From spectral theorem:

$$(X^T X)\phi = \lambda\phi \implies (X^T X)\phi - \lambda I\phi = 0$$

$$((X^T X) - \lambda I)\phi = 0$$

$$\begin{bmatrix} 30 - \lambda & 28 \\ 28 & 30 - \lambda \end{bmatrix} = 0 \implies \lambda = 58 \text{ and } \lambda = 2$$

EXAMPLE - PCA

$$X = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix}$$

From spectral theorem:

$$(X^T X)\phi = \lambda\phi$$

$$\begin{bmatrix} 30 & 28 \\ 28 & 30 \end{bmatrix} \begin{bmatrix} \phi_{11} \\ \phi_{12} \end{bmatrix} = 58 \begin{bmatrix} \phi_{11} \\ \phi_{12} \end{bmatrix} \implies \phi_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

EXAMPLE - PCA

$$X = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix}$$

From spectral theorem: $(X^T X)\phi = \lambda\phi$

$$\begin{bmatrix} 30 & 28 \\ 28 & 30 \end{bmatrix} \begin{bmatrix} \phi_{11} \\ \phi_{12} \end{bmatrix} = 58 \begin{bmatrix} \phi_{11} \\ \phi_{12} \end{bmatrix} \Rightarrow \phi_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\begin{bmatrix} 30 & 28 \\ 28 & 30 \end{bmatrix} \begin{bmatrix} \phi_{21} \\ \phi_{22} \end{bmatrix} = 2 \begin{bmatrix} \phi_{21} \\ \phi_{22} \end{bmatrix} \Rightarrow \phi_2 = \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

EXAMPLE - PCA

$$X = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix}$$

From spectral theorem: $(X^T X)\phi = \lambda\phi$

$$\phi_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \quad \lambda_1 = 58$$

$$\phi_2 = \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \quad \lambda_2 = 2$$

$$\phi = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

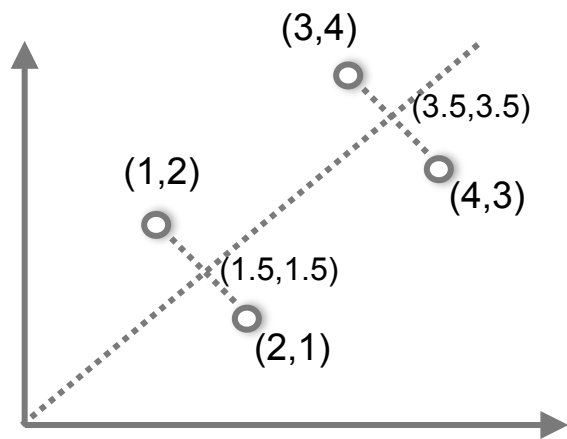
EXAMPLE - PCA

$$X = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} \quad \phi_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \quad \lambda_1 = 58 \quad \phi_2 = \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \quad \lambda_2 = 2$$

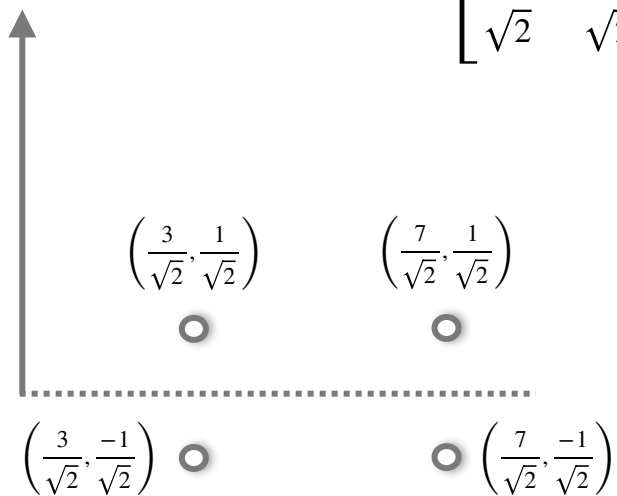
$$Z = X\phi = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{3}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{3}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{7}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{7}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}$$

EXAMPLE - PCA

$$X = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix}$$

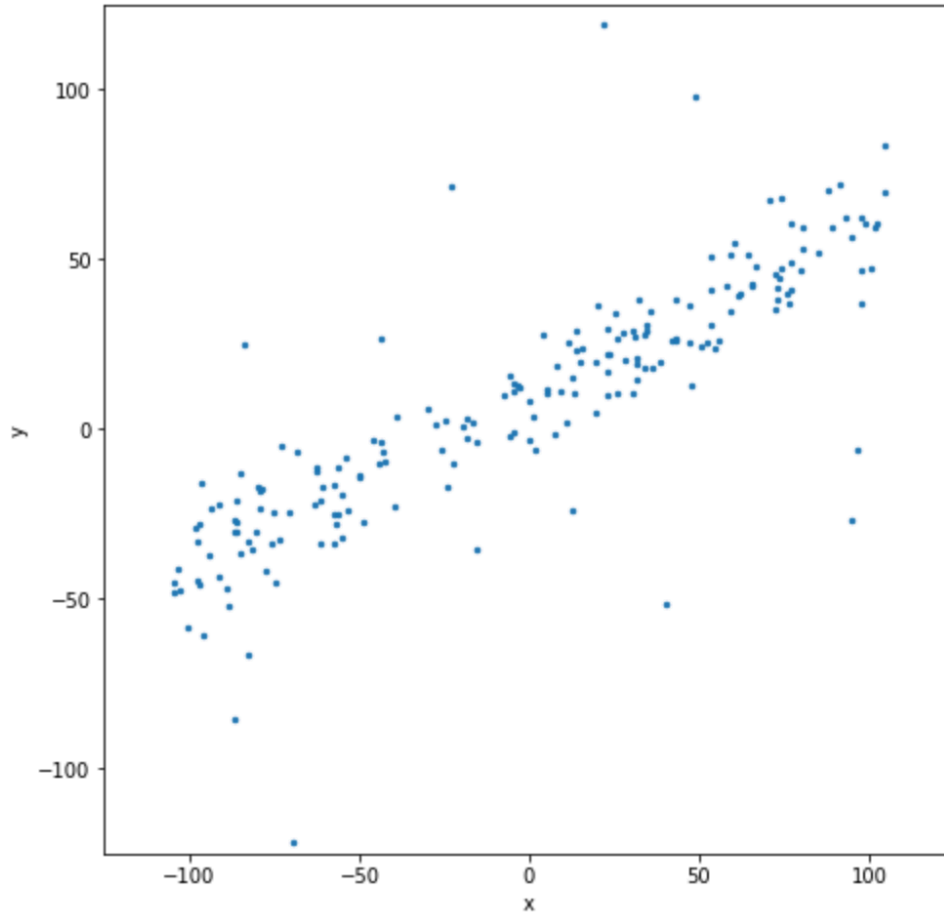


$$Z = \begin{bmatrix} \frac{3}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{3}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{7}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{7}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}$$



PCA STEPS -

STEP 1 MEAN SUBTRACTION



PCA STEPS -

STEP 2 COVARIANCE MATRIX

```
array([[ -84.07963403, -29.20401342],  
       [-94.40799878, -24.21822549],  
       [-20.0794366 , -18.24247399],  
       [ 62.63769935,  46.8234735 ]])
```

top 5 rows of mean centered data

```
array([[3881.30854873, 1897.16347756],  
       [1897.16347756, 1557.15527689]])
```

covariance matrix

PCA STEPS -

STEP 3 EIGEN VALUES & EIGEN VECTORS OF COVARIANCE MATRIX

```
eig_vals, eig_vecs = np.linalg.eig(cov)
```

```
eig_vals
```

```
array([4944.01310842,  494.4507172  ])
```

```
eig_vecs
```

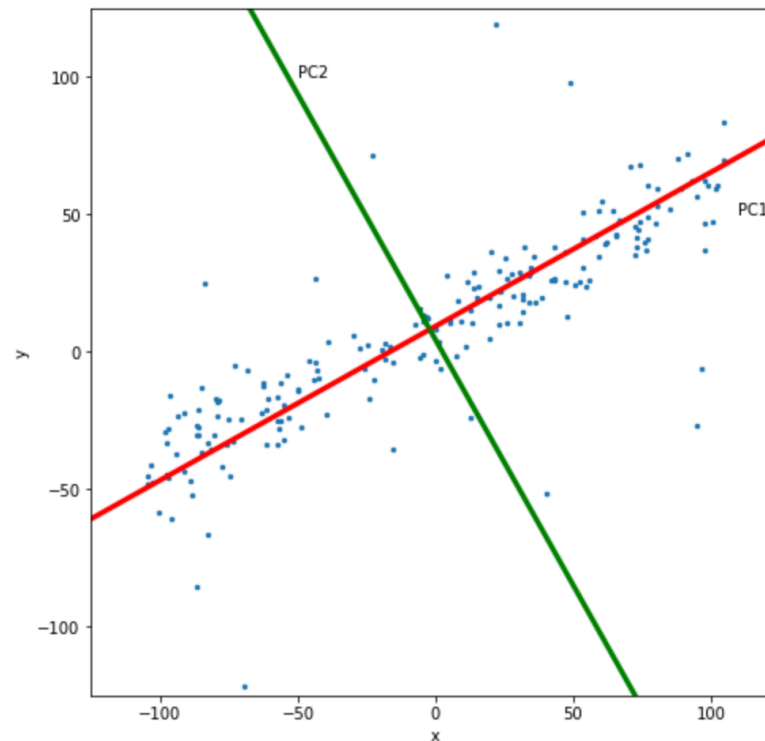
```
array([[ 0.87244857, -0.48870594],  
       [ 0.48870594,  0.87244857]])
```

PCA STEPS -

STEP 4 - PRINCIPAL COMPONENTS

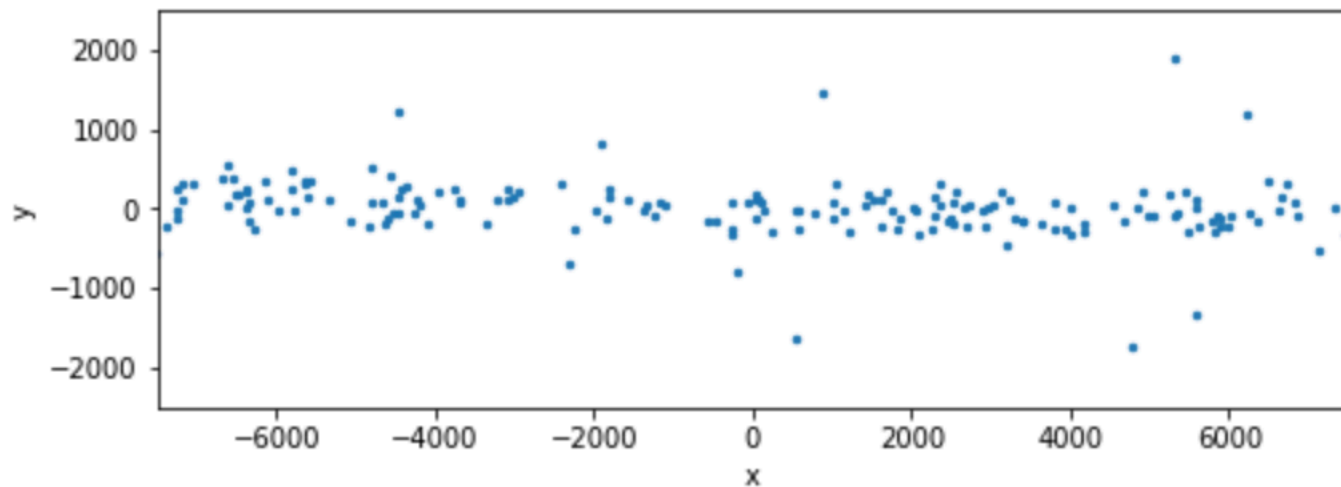
Multiply each eigen vector by its corresponding eigen value (usually square root)

Plot them on top of the data



PCA STEPS -

STEP 5 - PROJECT DATA ALONG DOMINANT PC



$$newData = PC1 \times oldData$$

HOW MANY PRINCIPAL COMPONENTS ?

- How many PCs should we consider in post-hoc analysis?
- One result of PCA is a measure of the variance to each PC relative to the total variance of the dataset.
- We can calculate the percentage of variance explained for the m-th PC:

$$PVE_m = \frac{\sum_{i=1}^n z_{im}^2}{\sum_{j=1}^p \sum_{i=1}^n x_{ij}^2}$$

HOW MANY PRINCIPAL COMPONENTS ?

- We can calculate the percentage of variance explained for the m-th PC:

$$PVE_m = \frac{\sum_{i=1}^n z_{im}^2}{\sum_{j=1}^p \sum_{i=1}^n x_{ij}^2}$$

