Tutorial 3: K-means and PCA

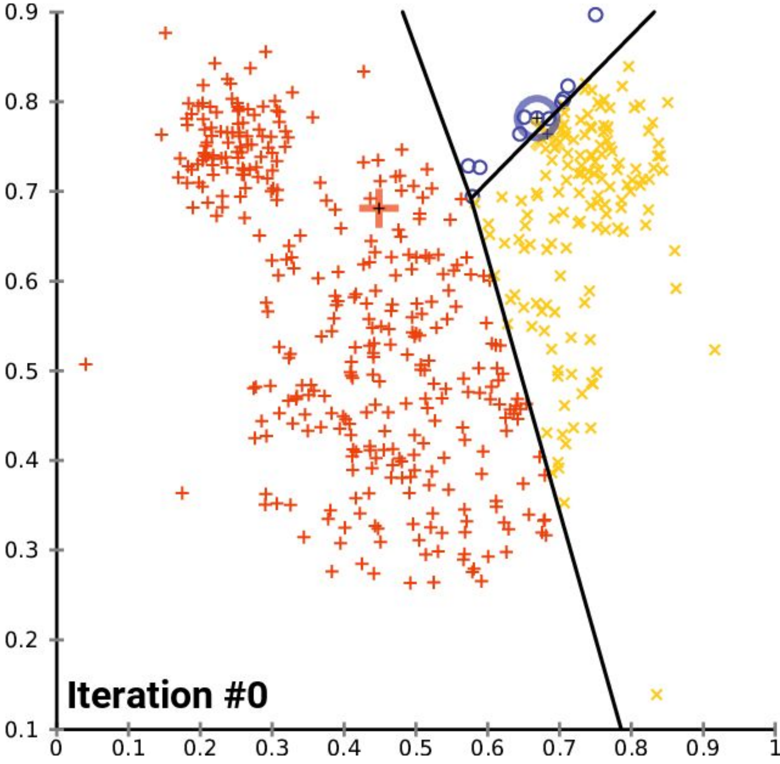
Wei Mao

Content

- K-means
  - K-means
  - K-means ++
- Principal Component Analysis (PCA)
  - Eigenvalue Decomposition
  - PCA
- Assignment 2 of ENGN6528

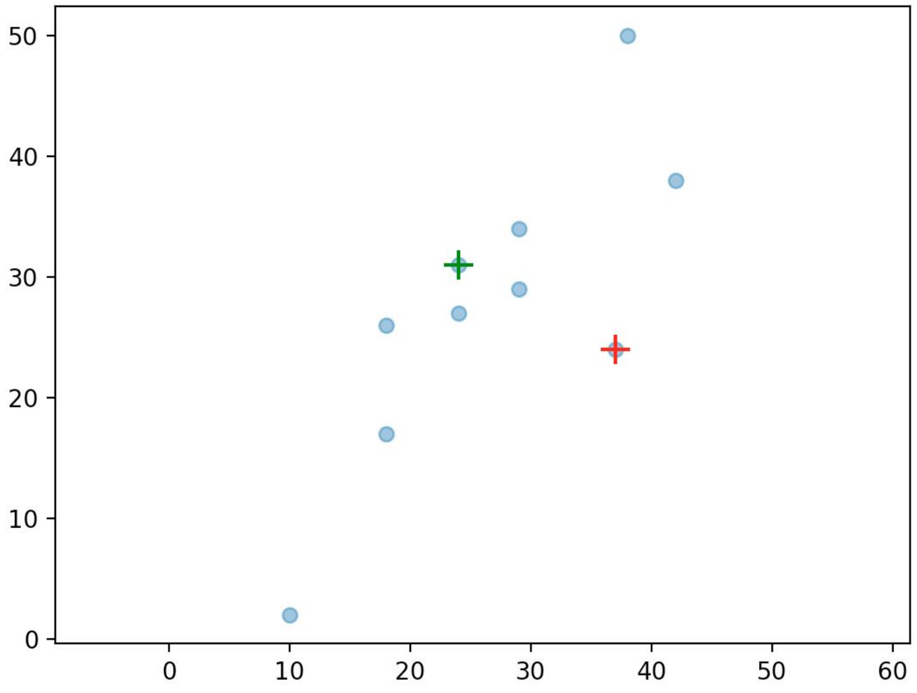
K-means

1. Randomly choose K datapoints as cluster centres
2. Assign datapoints to one of the cluster based on the distance to the centre
3. Update the centre of each cluster by taking the mean over all datapoints in this cluster.
4. Repeat 2 and 3 until convergence.
    a. Here convergence can be defined as the distance between new centre and old centre are small enough



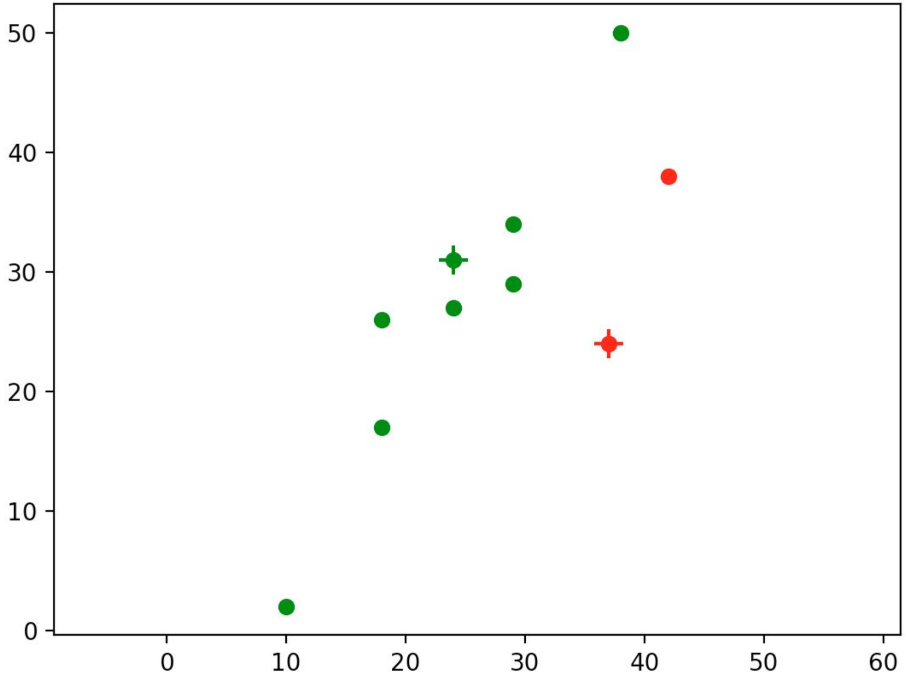
K-means-step by step

1. Randomly choose K datapoints as cluster centres



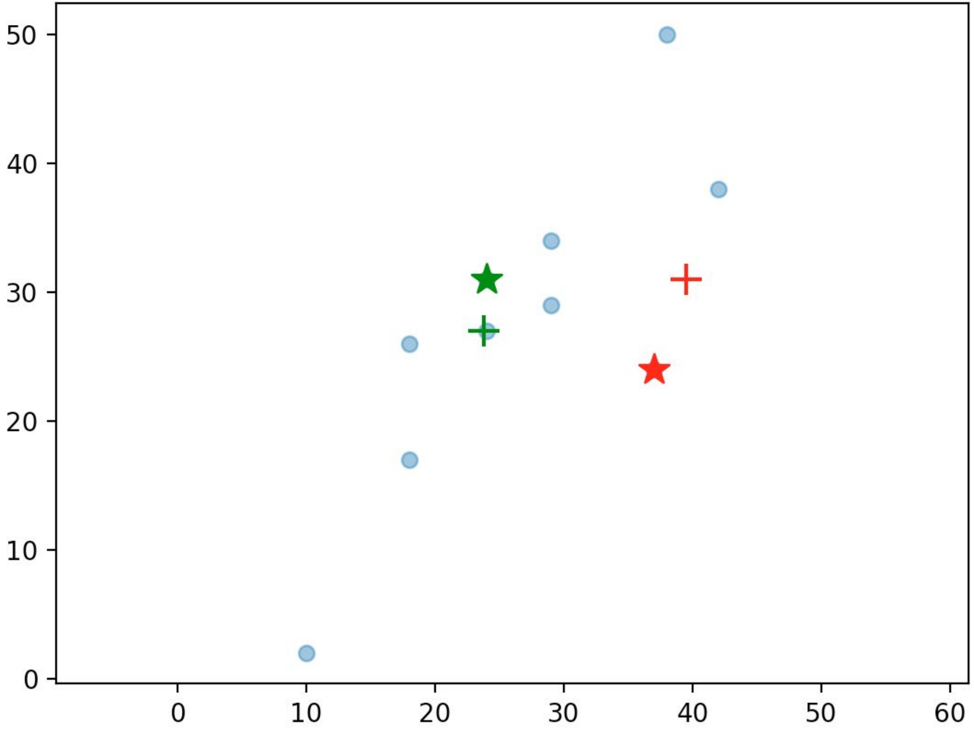
K-means

1. Randomly choose K datapoints as cluster centres
2. Assign datapoints to one of the cluster based on the distance to the centre



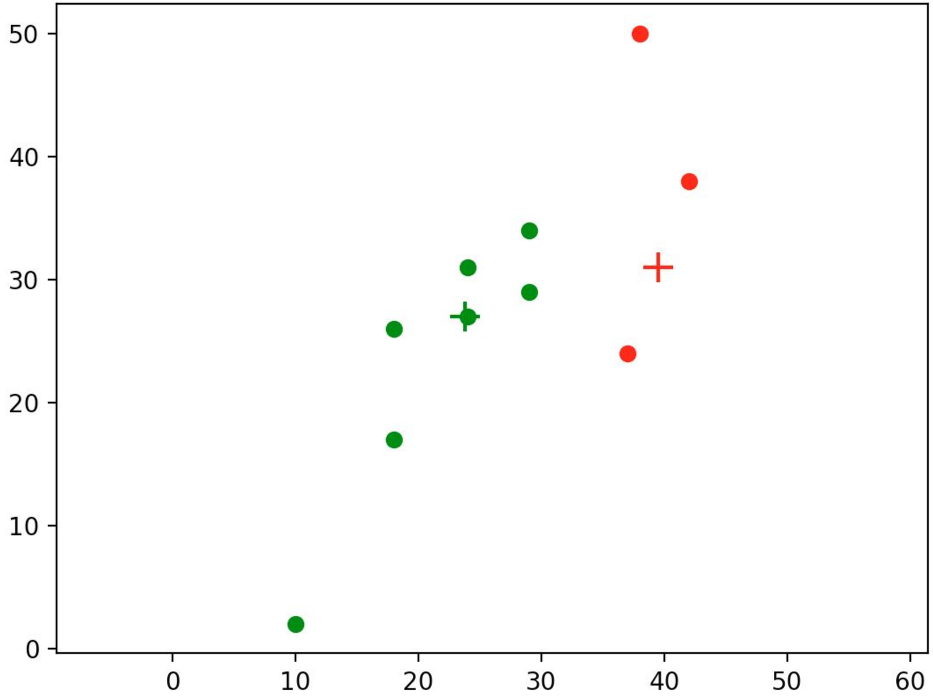
K-means

1. Randomly choose K datapoints as cluster centres
2. Assign datapoints to one of the cluster based on the distance to the centre
3. Update the centre of each cluster by taking the mean over all datapoints in this cluster.



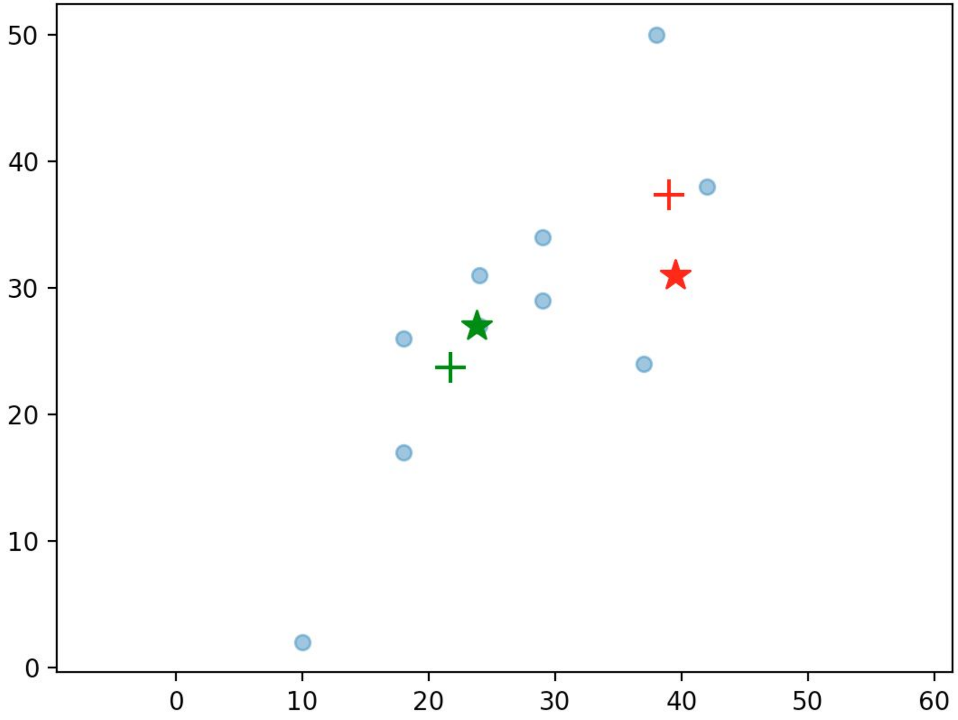
K-means

1. Randomly choose K datapoints as cluster centres
2. Assign datapoints to one of the cluster based on the distance to the centre
3. Update the centre of each cluster by taking the mean over all datapoints in this cluster.
4. Repeat 2 and 3 until convergence.



K-means

1. Randomly choose K datapoints as cluster centres
2. Assign datapoints to one of the cluster based on the distance to the centre
3. Update the centre of each cluster by taking the mean over all datapoints in this cluster.
4. Repeat 2 and 3 until convergence.



K-means ++

Obtaining good initialization for faster convergence and better performance.

Steps:

1. Randomly choose one data point as the first cluster centre.
2. For each data point **x**, compute the minimal distance d(**x**) from **x** to one of the cluster centres chosen.
3. Randomly choose the next cluster centre from remaining data points with the probability proportional to the squared distance computed in step 2
4. Repeat step 2 and 3 until find K cluster centroids.
5. Performing the standard K-means algorithm with the initial centres.

Eigenvalue Decomposition-Eigenvector

- **Eigenvector**
  - An eigenvector of a **square matrix** is a vector that do not change its direction under the linear transformation defined by the matrix.

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

eigenvector eigenvalue

Eigenvalue Decomposition-Eigenvector

- **How to find the eigenvectors and eigenvalues?**

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{I}\mathbf{v}$$

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = 0$$

Equivalent to first solve $det(\mathbf{A} - \lambda\mathbf{I}) = 0$ for λ

And, given λ to solve $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = 0$ for $\mathbf{v}$

Eigenvalue Decomposition-Eigenvector

- **E1. find the eigenvectors and eigenvalues of**

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$det(\mathbf{A} - \lambda\mathbf{I}) = 0$$

$$\begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = 0$$

$$(\lambda - 1)(\lambda - 3) = 0$$

$$=> \lambda_1 = 1, \lambda_2 = 3$$

Eigenvalue Decomposition-Eigenvector

- **E1. find the eigenvectors and eigenvalues of**

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

When $\lambda_1 = 1$,

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = 0$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \mathbf{v}_1 = 0$$

solve it and get $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

When $\lambda_1 = 3$,

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = 0$$

$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \mathbf{v}_2 = 0$$

solve it and get $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Eigenvalue Decomposition

- If a square matrix $\mathbf{A} \in \mathbb{R}^{m \times m}$ has m independent eigenvectors, then it can be factorized as

$$\mathbf{A} = \mathbf{U}\Lambda\mathbf{U}^{-1}$$

Where each column of $\mathbf{U}$ is an eigenvector and the corresponding diagonal

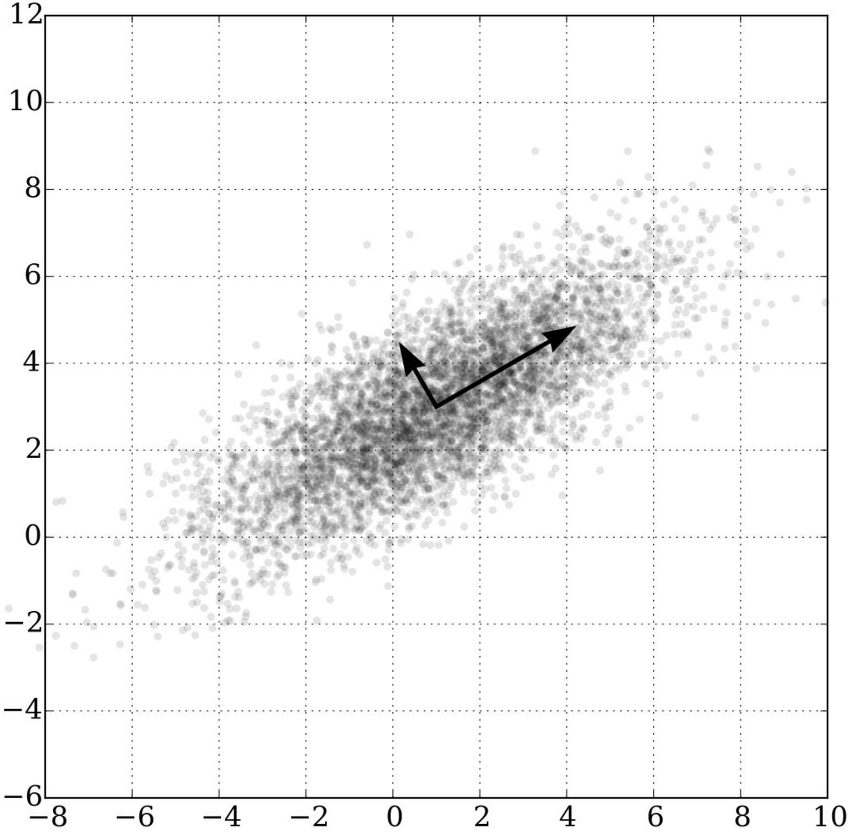
Elements of Λ is the corresponding eigenvalue.

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} * \begin{bmatrix} 0.5 & -0.5 \\ 0.5 & 0.5 \end{bmatrix}$$

Principal Component Analysis

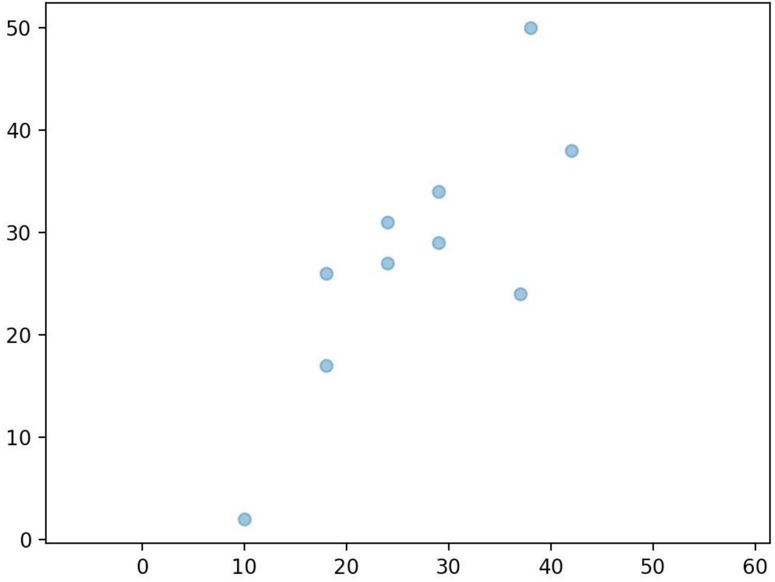
- Given a set of data points, the eigenvectors of their covariance matrix defines the principal components of those data points.
- Often used for dimension deduction.



Principal Component Analysis-step by step

- Subtract the datapoints by their mean
- Compute the covariance matrix
- Compute the eigenvectors and eigenvalues of the covariance matrix
- Rank the eigenvectors by their eigenvalues in descending order
- Obtain the top-k eigenvectors and then project datapoints to the new space defined by those eigenvectors

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
x	37	24	29	42	38	10	29	18	18	24
y	24	27	34	38	50	2	29	17	26	31

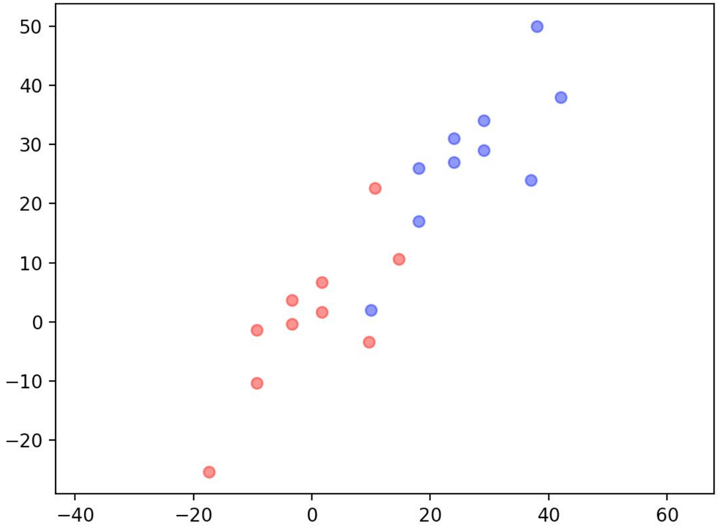


Principal Component Analysis-step by step

- Subtract the datapoints by their mean

$$\hat{\mathbf{X}} = \mathbf{X} - \overline{\mathbf{X}}$$

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
x	10.1	-2.9	2.1	15.1	11.1	-16.9	2.1	-8.9	-8.9	-2.9
y	-3.8	-0.8	6.2	10.2	22.2	-25.8	1.2	-10.8	-1.8	3.2



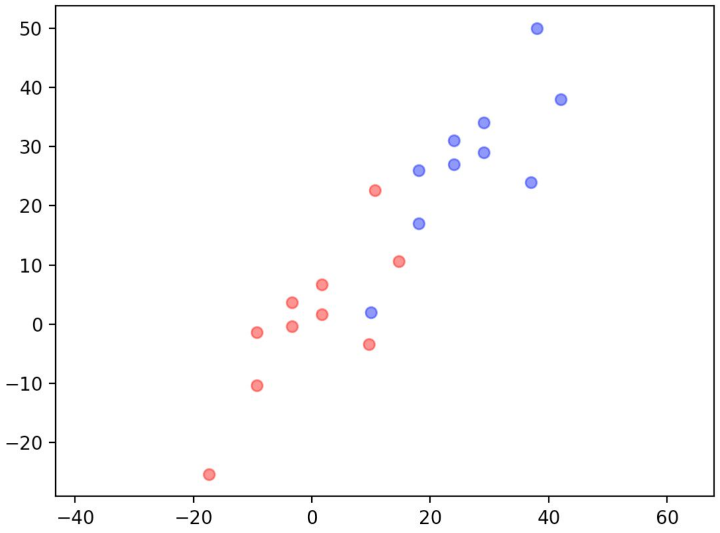
Principal Component Analysis-step by step

- Compute the covariance matrix

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
x	10.1	-2.9	2.1	15.1	11.1	-16.9	2.1	-8.9	-8.9	-2.9
y	-3.8	-0.8	6.2	10.2	22.2	-25.8	1.2	-10.8	-1.8	3.2

$$\mathbf{C} = \frac{1}{N}\hat{\mathbf{X}}^T\hat{\mathbf{X}}$$

$$= \begin{bmatrix} 92.3 & 91.9 \\ 91.9 & 144.8 \end{bmatrix}$$



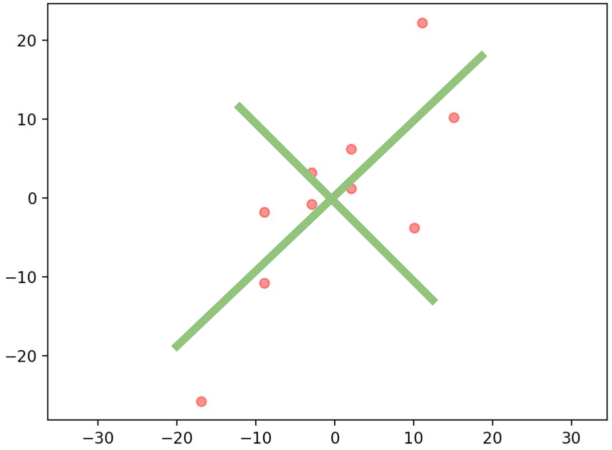
Principal Component Analysis-step by step

- Compute the eigenvectors and eigenvalues of the covariance matrix
- Rank the eigenvectors by their eigenvalues in descending order

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
x	10.1	-2.9	2.1	15.1	11.1	-16.9	2.1	-8.9	-8.9	-2.9
y	-3.8	-0.8	6.2	10.2	22.2	-25.8	1.2	-10.8	-1.8	3.2

$$\lambda = [214.1, 23.0]$$

$$\mathbf{v} = \begin{bmatrix} -0.6 & -0.8 \\ -0.8 & 0.6 \end{bmatrix}$$



Principal Component Analysis-step by step

- Obtain the top-k eigenvectors and then project datapoints to the new space defined by those eigenvectors

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
x	10.1	-2.9	2.1	15.1	11.1	-16.9	2.1	-8.9	-8.9	-2.9
y	-3.8	-0.8	6.2	10.2	22.2	-25.8	1.2	-10.8	-1.8	3.2