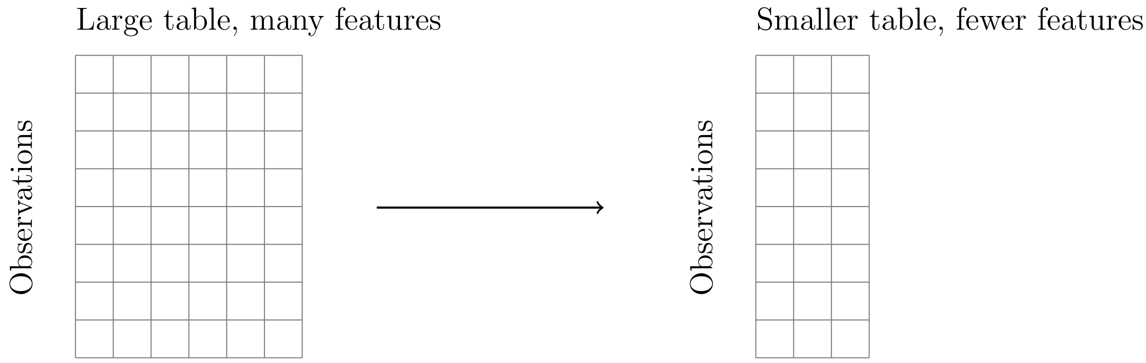
# Notes on Dimensionality Reduction

## Dimensionality Reduction

Dimensionality reduction involves reducing the number of features (columns) in a dataset while retaining as much information as possible. This process is crucial for simplifying models, visualizing high-dimensional data, and improving computational efficiency.




## Projections

Projections are fundamental operations in linear algebra used to map points onto a subspace. Here, we illustrate an example where points in two-dimensional space are projected onto a line defined by the vector $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

### Example: Projection onto a Line

Given the points:

$$\begin{bmatrix} 1.0 & 1.0 \\ 1.2 & 1.6 \\ -0.5 & 0.2 \\ -1.3 & -0.6 \end{bmatrix},$$

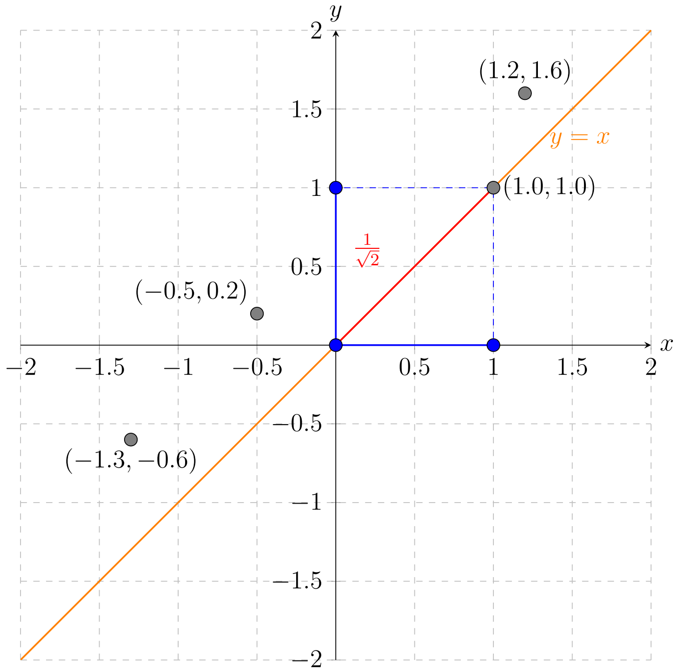
we aim to project them onto the line $y = x$ using the unit vector:

$$\mathbf{u} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The projection formula is:

$$\text{Proj}_{\mathbf{u}}(\mathbf{x}) = (\mathbf{x} \cdot \mathbf{u})\mathbf{u},$$

where $\mathbf{x} \cdot \mathbf{u}$ is the dot product.




$$\text{Projection} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \left( \frac{x+y}{\sqrt{2}} \right) = \frac{x+y}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{x+y}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Reinterpret the coefficient:

$$\text{Projection} = \frac{x+y}{\sqrt{2}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Final Step: Simplify using the unit vector definition:

$$\text{Projection} = \frac{x+y}{\sqrt{2}}\mathbf{v}, \quad \text{where } \mathbf{v} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

1. **Point:** $(1.0, 1.0)$

$$\frac{x+y}{\sqrt{2}} = \frac{1.0+1.0}{\sqrt{2}} = \frac{2}{\sqrt{2}} = \sqrt{2}.$$

2. **Point:** $(1.2, 1.6)$

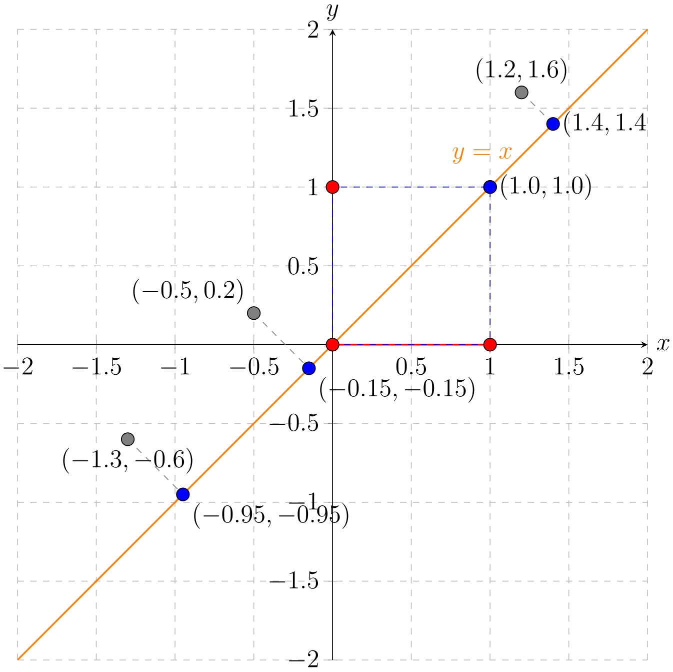
$$\frac{x+y}{\sqrt{2}} = \frac{1.2+1.6}{\sqrt{2}} = \frac{2.8}{\sqrt{2}} = 1.9799.$$

3. **Point:** $(-0.5, 0.2)$

$$\frac{x+y}{\sqrt{2}} = \frac{-0.5+0.2}{\sqrt{2}} = \frac{-0.3}{\sqrt{2}} = -0.2121.$$

4. **Point:** $(-1.3, -0.6)$

$$\frac{x+y}{\sqrt{2}} = \frac{-1.3+(-0.6)}{\sqrt{2}} = \frac{-1.9}{\sqrt{2}} = -1.344.$$




# Projections of Matrices

- The projection normalizes the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ to ensure the correct scaling of the projected matrix.
- When projecting onto multiple vectors, $\mathbf{V}$ acts as the transformation matrix containing the directions of the projections.
- This technique is commonly used in dimensionality reduction, such as Principal Component Analysis (PCA), where data is projected onto a reduced set of basis vectors.

## Projection onto a Single Vector

To project a matrix $A$ onto a vector $\mathbf{v}$, we use the formula:

$$A_P = A\frac{\mathbf{v}}{\|\mathbf{v}\|_2},$$

where:
- $A$ is the original matrix of size $r \times c$,
- $\mathbf{v}$ is the vector of size $c \times 1$,
- $\|\mathbf{v}\|_2$ is the Euclidean norm (magnitude) of $\mathbf{v}$,
- $A_P$ is the resulting projection matrix of size $r \times 1$.

## Projection onto Multiple Vectors

To project a matrix $A$ onto two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, we use the formula:

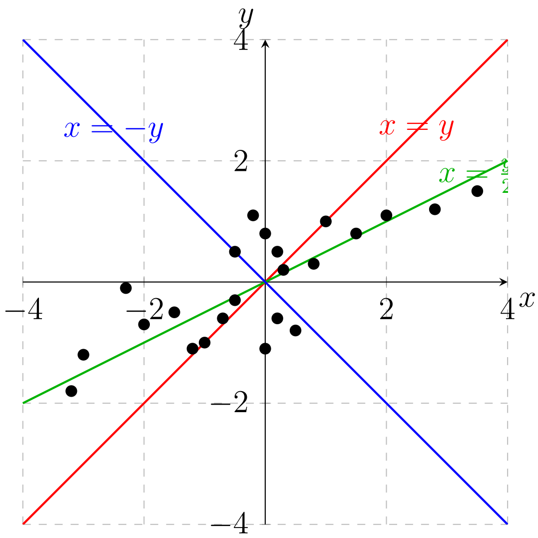
$$A_P = A \underbrace{\begin{bmatrix} \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|_2} & \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|_2} \end{bmatrix}}_{\mathbf{V}},$$

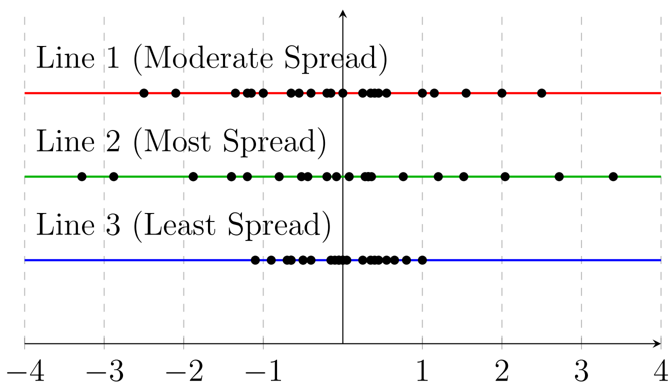
where:
- $A$ is the original matrix of size $r \times c$,
- $\mathbf{v}_1, \mathbf{v}_2$ are the projection vectors, each of size $c \times 1$,
- $\mathbf{V}$ is the matrix of normalized projection vectors, of size $c \times 2$,
- $A_P$ is the resulting projection matrix of size $r \times 2$.

# Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a statistical technique used to reduce the dimensionality of data by transforming it into a new set of variables (principal components) that are uncorrelated and ordered by the amount of variance they capture.

- PCA identifies directions (principal components) in which the data varies the most.
- The first principal component captures the most variance, followed by the second, and so on.
- The data is projected onto these principal components to reduce dimensions while preserving as much information as possible.




- The dataset consists of scattered points in two dimensions.
- Different lines (e.g., $x = y$, $x = -y$, $2x = y$) represent possible directions for projecting the data.
- The red line ($2x = y$) shows the direction with the most spread, hence capturing the most information (first principal component).
- The blue line ($x = -y$) has the least spread, capturing the least information.

# Mean of the Data

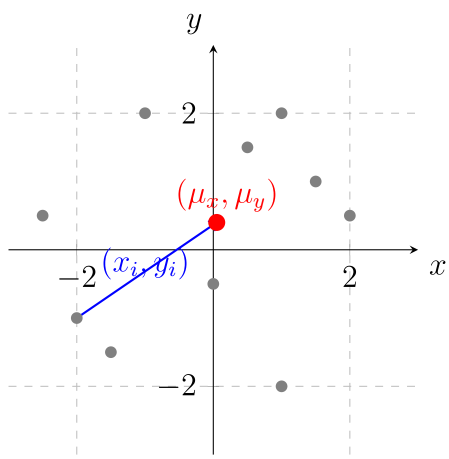
The mean is often referred to as "the average of the data." It represents the central point of a dataset and is computed as the arithmetic mean of all data points in each dimension.

## Definition

Given a set of data points $(x_i, y_i)$, the mean in each dimension is computed as:

$$\mu_x = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \mu_y = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

where $n$ is the number of data points.
The mean is represented by the point $(\mu_x, \mu_y)$, which serves as the center of mass of the data points. In the illustration, the red point represents the mean, and arrows point from a sample data point $(x_i, y_i)$ to the mean point.



# Variance of the Data

Variance quantifies the spread of the data points around the mean. It is computed as the average of the squared differences between each data point and the mean.

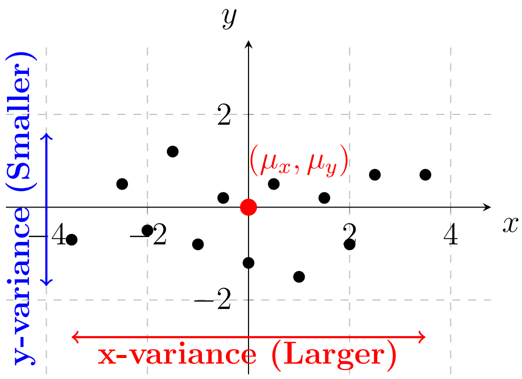
The variance in each dimension is defined as:

$$\mathrm{Var}(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu_x)^2,$$

$$\mathrm{Var}(y) = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \mu_y)^2,$$

where $n$ is the number of data points, and $(\mu_x, \mu_y)$ is the mean. Variance reflects how much the data points deviate from the mean. A larger variance indicates a wider spread of data points.

In the illustration, the x-dimension shows larger variance, while the y-dimension has smaller variance.




# Covariance of the Data

Covariance measures the direction of the linear relationship between two variables. It indicates how changes in one variable correspond to changes in another. A positive covariance means the variables increase together, a negative covariance indicates one increases while the other decreases, and a near-zero covariance suggests no linear relationship.

$$\mathrm{Cov}(x, y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y),$$

where:

- $n$ is the number of data points,
- $x_i$ and $y_i$ are the individual data points,
- $\mu_x$ and $\mu_y$ are the means of the $x$- and $y$-coordinates, respectively.

## Example Calculation

Suppose we have the following data points:

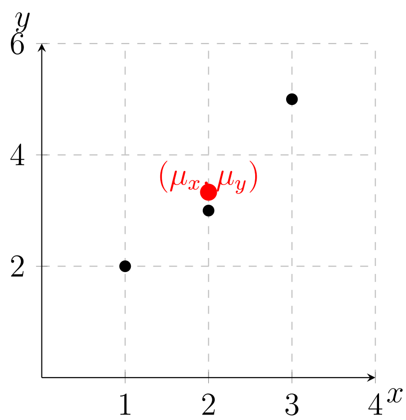
$$(x, y) = \{(1, 2), (2, 3), (3, 5)\}, \quad \mu_x = 2, \quad \mu_y = 3.33.$$

We calculate covariance as:

$$\mathrm{Cov}(x, y) = \frac{1}{3-1}[(1-2)(2-3.33) + (2-2)(3-3.33) \\ +(3-2)(5-3.33)]$$

$$= \frac{1}{2}(-1 \cdot -1.33 + 0 \cdot -0.33 + 1 \cdot 1.67) = 1.5.$$

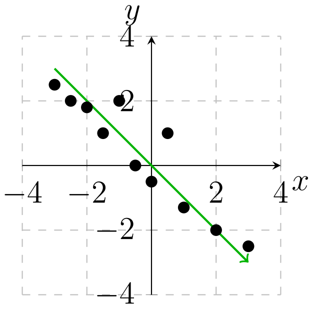
This positive covariance indicates that $x$ and $y$ increase together.



# Covariance Matrix

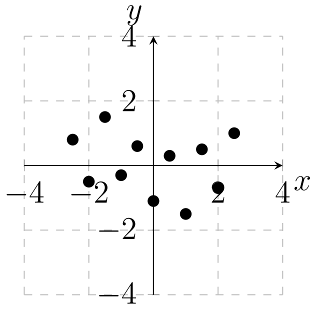
The covariance matrix provides a comprehensive way to understand the relationships between variables. It summarizes both the variances of each variable and their covariances with one another. The covariance matrix is given by:

$$\mathbf{C} = \begin{bmatrix} \mathrm{Var}(x) & \mathrm{Cov}(x,y) \\ \mathrm{Cov}(y,x) & \mathrm{Var}(y) \end{bmatrix}.$$



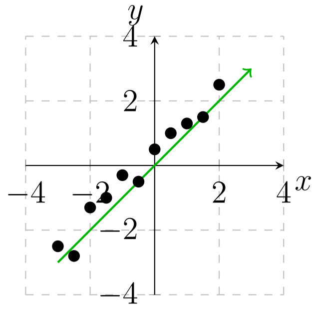

**Negative Covariance**
Indicates that as $x$ increases, $y$ tends to decrease

$$\mathbf{C} = \begin{bmatrix} 2.1 & -1.5 \\ -1.5 & 1.8 \end{bmatrix}$$




**Zero Covariance**
Indicates no linear relationship between $x$ and $y$

$$\mathbf{C} = \begin{bmatrix} 1.5 & 0.0 \\ 0.0 & 1.2 \end{bmatrix}$$




**Positive Covariance**
Indicates that as $x$ increases, $y$ tends to increase

$$\mathbf{C} = \begin{bmatrix} 2.2 & 1.5 \\ 1.5 & 2.8 \end{bmatrix}$$

# Covariance Matrix

The covariance matrix represents the pairwise covariances between multiple variables. It is a square matrix where the diagonal elements represent the variances of each variable, and the off-diagonal elements represent the covariances between variables.

$$\mathbf{C} = \frac{1}{n-1}(\mathbf{A} - \mu)^T(\mathbf{A} - \mu),$$

where:

- $\mathbf{A}$ is the matrix of data points, where each row is a data point $[x_i, y_i]$,
- $\mu$ is the matrix of mean values repeated for each data point,
- $n$ is the number of data points.

## Step 1: Arrange the Data and Subtract the Mean

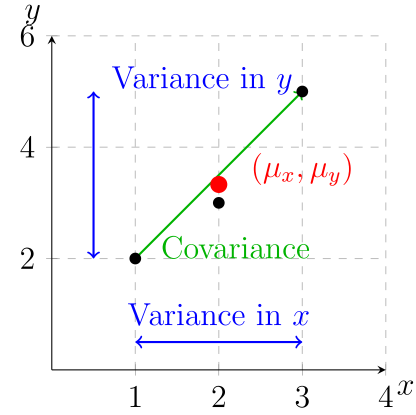
For the dataset $(x, y) = \{(1,2), (2,3), (3,5)\}$, arrange data with a different feature in each column and calculate the average of each column.

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 5 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_x & \mu_y \\ \mu_x & \mu_y \\ \mu_x & \mu_y \end{bmatrix} = \begin{bmatrix} 2 & 3.33 \\ 2 & 3.33 \\ 2 & 3.33 \end{bmatrix}.$$

Subtract the averages $\mu$ from respective columns of $\mathbf{A}$:

$$\mathbf{A} - \mu = \begin{bmatrix} 1-2 & 2-3.33 \\ 2-2 & 3-3.33 \\ 3-2 & 5-3.33 \end{bmatrix} = \begin{bmatrix} -1 & -1.33 \\ 0 & -0.33 \\ 1 & 1.67 \end{bmatrix}.$$

### Step 2: Compute the Covariance Matrix




The covariance matrix is computed as:

$$\mathbf{C} = \frac{1}{n-1}(\mathbf{A}-\mu)^T(\mathbf{A}-\mu).$$

First, compute$(\mathbf{A}-\mu)$ and $(\mathbf{A}-\mu)^T$:

$$(\mathbf{A}-\mu) = \begin{bmatrix} -1 & -1.33 \\ 0 & -0.33 \\ 1 & 1.67 \end{bmatrix}, \quad (\mathbf{A}-\mu)^T = \begin{bmatrix} -1 & 0 & 1 \\ -1.33 & -0.33 & 1.67 \end{bmatrix}.$$

Next, multiply the matrices:

$$(\mathbf{A}-\mu)^T(\mathbf{A}-\mu) = \begin{bmatrix} -1 & 0 & 1 \\ -1.33 & -0.33 & 1.67 \end{bmatrix} \begin{bmatrix} -1 & -1.33 \\ 0 & -0.33 \\ 1 & 1.67 \end{bmatrix} = \begin{bmatrix} 2 & 2.5 \\ 2.5 & 3.11 \end{bmatrix}.$$

Finally, divide by $n-1=2$. The covariance matrix is:

$$\mathbf{C} = \frac{1}{2}\begin{bmatrix} 2 & 2.5 \\ 2.5 & 3.11 \end{bmatrix} = \begin{bmatrix} 1 & 1.25 \\ 1.25 & 1.555 \end{bmatrix} = \begin{bmatrix} \mathrm{Var}(x) & \mathrm{Cov}(x,y) \\ \mathrm{Cov}(y,x) & \mathrm{Var}(y) \end{bmatrix}.$$

This shows the variances along the diagonal and the covariances off-diagonal.

## Mathematical Formulation of PCA

Principal Component Analysis (PCA) is a statistical method used to reduce the dimensionality of data while preserving as much variance as possible.
- The Principal Components (PC) represent the directions of maximum variance in the data after dimensionality reduction.
- The projections of the data onto v are the new coordinates of the data in the reduced space.

### 1. Create the Data Matrix $X$

We begin with $n$ observations of $p$ variables. These variables are arranged in a matrix $X$ of size $n \times p$:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

### 2. Center the Data

Subtract the mean from each variable ($\mu_j$ is the mean of the $j$-th variable.):

$$X - \mu = \begin{bmatrix} x_{11}-\mu_1 & x_{12}-\mu_2 & \cdots & x_{1p}-\mu_p \\ x_{21}-\mu_1 & x_{22}-\mu_2 & \cdots & x_{2p}-\mu_p \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1}-\mu_1 & x_{n2}-\mu_2 & \cdots & x_{np}-\mu_p \end{bmatrix}$$

## 3. Calculate the Covariance Matrix

The covariance matrix $C$ is computed as:

$$C = \frac{1}{n-1}(X-\mu)^\top(X-\mu)$$

The covariance matrix $C$ is a $p \times p$ symmetric matrix:

$$C = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_p) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \cdots & \text{Cov}(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_p, X_1) & \text{Cov}(X_p, X_2) & \cdots & \text{Var}(X_p) \end{bmatrix}$$

## 4. Compute Eigenvalues and Eigenvectors

Find the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ and their corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ of the covariance matrix $C$:

$$C\mathbf{v}_i = \lambda_i \mathbf{v}_i$$

The eigenvalues represent the variance explained by each principal component.

## 5. Create the Projection Matrix

Choose the top $k$ eigenvectors (corresponding to the largest $k$ eigenvalues). These eigenvectors form the projection matrix $V$:

$$V = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \end{bmatrix}$$

## 6. Project the Data onto the New Basis

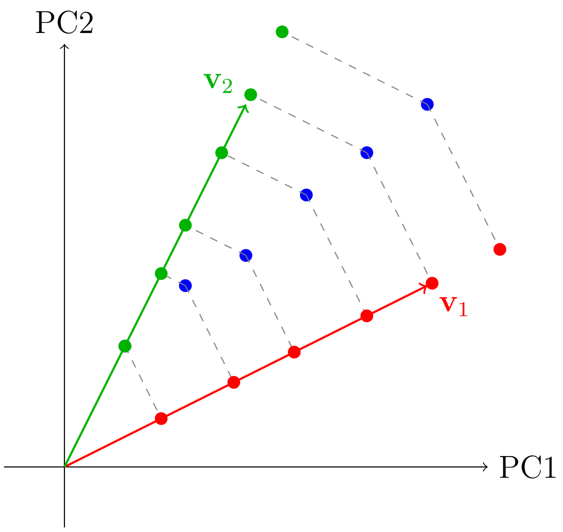
The centered data is projected onto the new basis defined by $V$:

$$X_{PCA} = (X-\mu)V$$

Here, $X_{PCA}$ is the reduced-dimensional representation of the data.

## Example: Reducing 5 Variables to 2 Principal Components $\mathbb{R}^5$ to $\mathbb{R}^2$:

1. Start with a dataset $X$ with $n$ observations and 5 variables.
2. Center the data by subtracting the mean of each variable.
3. Compute the $5 \times 5$ covariance matrix $C$.
4. Find the eigenvalues $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5$ and their corresponding eigenvectors.
5. Choose the 2 largest eigenvalues and their eigenvectors to form the projection matrix $V$.
6. Project the centered data $X - \mu$ onto $V$ to obtain $X_{PCA}$.



# Visual Understanding of PCA

Principal Component Analysis (PCA) is a method to identify the directions (principal components) along which data varies the most. These directions are given by the eigenvectors of the covariance matrix, and the magnitude of variation is given by the eigenvalues.

## Steps to Perform PCA

### 1. Compute the covariance matrix $C$ of the data

$$C = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i-\mu)(\mathbf{x}_i-\mu)^T = \begin{bmatrix} \mathrm{Var}(x) & \mathrm{Cov}(x,y) \\ \mathrm{Cov}(x,y) & \mathrm{Var}(y) \end{bmatrix} = \begin{bmatrix} 9 & 4 \\ 4 & 3 \end{bmatrix}.$$

where $\mu$ is the mean vector of the data points, and $\mathbf{x}_i$ are the data points. The covariance matrix $C$ quantifies the spread of the data.

### 2. Find the eigenvalues and eigenvectors of $C$

Eigenvalues $\lambda$ and eigenvectors $\mathbf{v}$ of $C$ satisfy the equation:

$$C\mathbf{v} = \lambda\mathbf{v},$$

To find the eigenvalues, solve the characteristic equation:

$$\det(C-\lambda I) = 0,$$

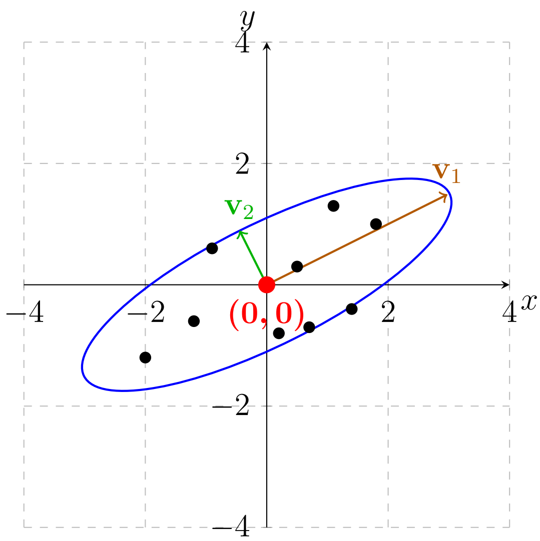
where $I$ is the identity matrix. Substituting $C = \begin{bmatrix} 9 & 4 \\ 4 & 3 \end{bmatrix}$:

$$\det\begin{bmatrix} 9-\lambda & 4 \\ 4 & 3-\lambda \end{bmatrix} = (9-\lambda)(3-\lambda)-4^2 = \lambda^2-12\lambda+11 = 0.$$

Solving this quadratic equation:

$$\lambda_1 = 11, \quad \lambda_2 = 1.$$

Next, find the eigenvectors by solving $(C-\lambda I)\mathbf{v} = 0$ for each eigenvalue.




$$\text{For } \lambda_1 = 11\text{: } \begin{bmatrix} -2 & 4 \\ 4 & -8 \end{bmatrix}\mathbf{v}_1 = 0 \quad \Longrightarrow \quad \mathbf{v}_1 = \begin{bmatrix} 0.8944 \\ 0.4472 \end{bmatrix}.$$

$$\text{For } \lambda_2 = 1\text{: } \begin{bmatrix} 8 & 4 \\ 4 & 2 \end{bmatrix}\mathbf{v}_2 = 0 \quad \Longrightarrow \quad \mathbf{v}_2 = \begin{bmatrix} -0.4472 \\ 0.8944 \end{bmatrix}.$$

### 3. Interpret the eigenvalues and eigenvectors

- The ellipse representing the spread of the data, with semi-axes lengths proportional to $\sqrt{\lambda_1}$ and $\sqrt{\lambda_2}$.
- The eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ represent the principal directions of the data. These directions are the axes of the ellipse that fits the data's spread.
- The eigenvalues $\lambda_1$ and $\lambda_2$ correspond to the variance along these directions.

The square roots of the eigenvalues give the lengths of the semi-axes of the ellipse:

$$\text{Length along } \mathbf{v}_1 = \sqrt{\lambda_1} = \sqrt{11} \approx 3.32, \quad \text{Length along } \mathbf{v}_2 = \sqrt{\lambda_2} = \sqrt{1} = 1.$$