Definition

Given a vector

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \in \mathbb{R}^N$$

the *mean* of X is the number

$$m_X = \frac{x_1 + \ldots + x_N}{N}$$

Notation. For a vector $X \in \mathbb{R}^N$ as above, by $\widetilde{X}$ we will denote the vector

$$\widetilde{X} = \begin{bmatrix} x_1 - m_X \\ \vdots \\ x_N - m_X \end{bmatrix}$$

We will say that $\widetilde{X}$ is the *demeaning* of X.

Note. $m_{\widetilde{X}} = 0$.

Example.

$X = \begin{bmatrix} 62 & 90 & 79 & 70 & 55 \end{bmatrix}^T$, $m_X = 71$

$\widetilde{X} = \begin{bmatrix} -9 & 19 & 8 & -1 & -16 \end{bmatrix}^T$, $m_{\widetilde{X}} = 0$

Definition

Given a vector

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \in \mathbb{R}^N$$

the *variance* of X is the number

$$\mathrm{Var}(X) = \frac{1}{N}\sum_{i=1}^{N}(x_i - m_X)^2$$

Note. $\mathrm{Var}(X) = \frac{1}{N}\widetilde{X}^T\widetilde{X}$.

$$\frac{1}{N}\widetilde{X}^T\widetilde{X} = \frac{1}{N}\widetilde{X}\bullet\widetilde{X} = \frac{1}{N}\sum_{i=1}^{N}(x_i - m_X)^2$$

dot product

Proposition

For a vector

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \in \mathbb{R}^N$$

we have

$$\mathrm{Var}(X) = \frac{1}{2N^2}\sum_{i,j}(x_i - x_j)^2$$

Definition

Given vectors

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \in \mathbb{R}^N$$

the *covariance* of X and Y is the number

$$\mathrm{Cov}(X, Y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - m_X)(y_i - m_Y)$$

Note.

1) $\mathrm{Cov}(X, Y) = \frac{1}{N}\widetilde{X}^T\widetilde{Y}$ $= \frac{1}{N}\widetilde{X}^T \bullet \widetilde{Y}$

dot product

2) $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$

Proposition

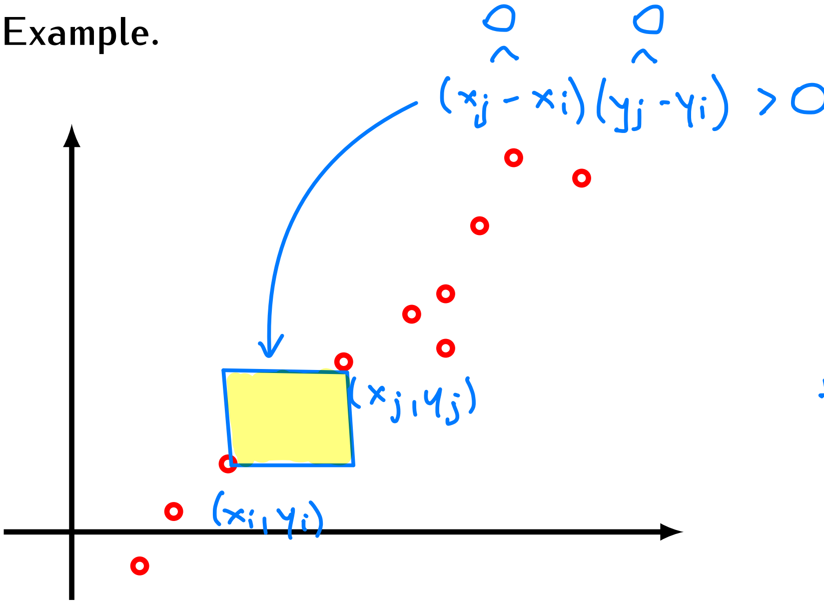
For vectors

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \in \mathbb{R}^N$$

we have

$$\mathrm{Cov}(X, Y) = \frac{1}{2N^2}\sum_{i,j}(x_i - x_j)(y_i - y_j)$$

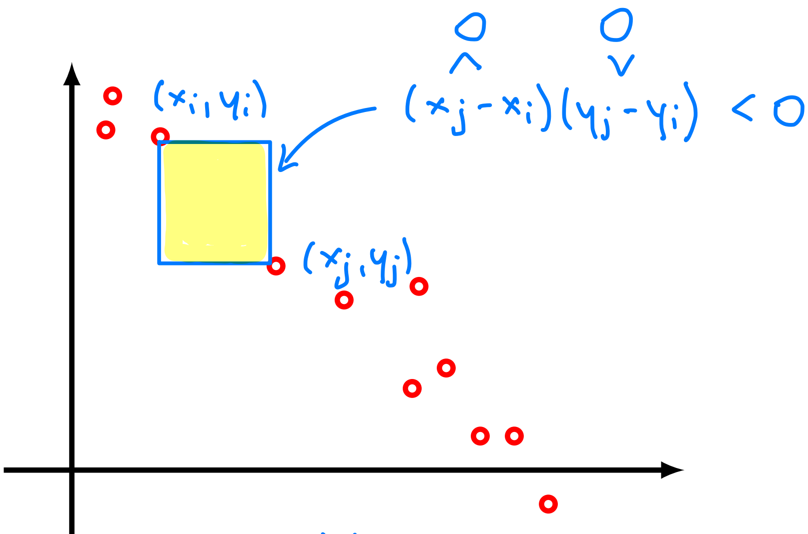
Example.



$\mathrm{Cov}(X, Y) > 0$

$\Downarrow$

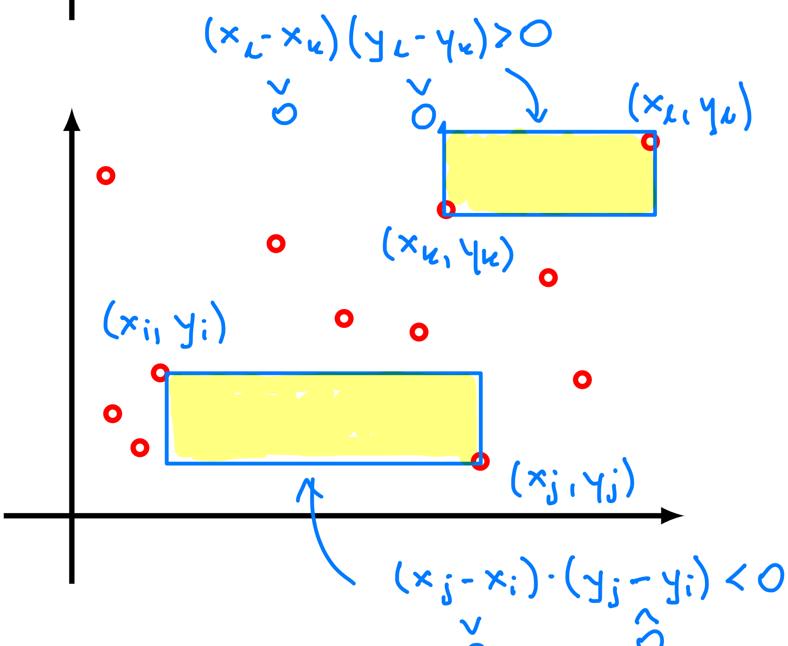
on average values of the vector Y increase when the corresponding values of the vector X increase



$\mathrm{Cov}(X, Y) < 0$

$\Downarrow$

on average values of the vector Y decrease when the corresponding values of the vector X increase



$\mathrm{Cov}(X, Y)$

$\Downarrow$

we cannot predict if a value of Y will increase or decrease if we know that the corresponding value of X increased