Example.

$$P = \begin{array}{r} \\ \text{Aly} \\ \text{Bob} \\ \text{Chen} \\ \text{Deb} \end{array} \begin{array}{c} \begin{array}{ccc} \text{height} & \text{weight} & \text{age} \end{array} \\ \begin{bmatrix} 62 & 141 & 19 \\ 82 & 164 & 21 \\ 79 & 154 & 19 \\ 70 & 135 & 25 \end{bmatrix} \end{array}$$

<u>**General form**</u>

$$A = \begin{bmatrix} \overbrace{\boxed{X_1} \quad \boxed{X_2} \quad \cdots \quad \boxed{X_M}}^{\text{feature vectors}} \end{bmatrix} = \left.\begin{bmatrix} \boxed{\quad S_1 \quad} \\ \boxed{\quad S_1 \quad} \\ \vdots \\ \boxed{\quad S_N \quad} \end{bmatrix}\right\} \text{sample vectors}$$

Notation. Given a data matrix $A = \begin{bmatrix} X_1 & X_2 & \dots & X_M \end{bmatrix}$ we will denote

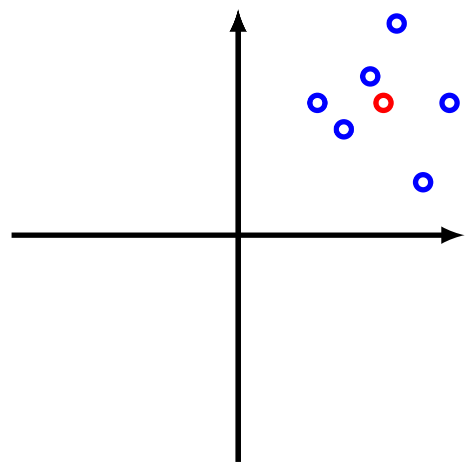
$$\widetilde{A} = \begin{bmatrix} \widetilde{X}_1 & \widetilde{X}_2 & \dots & \widetilde{X}_M \end{bmatrix}$$

where $\widetilde{X}_i$ is the demeaning of the vector X_i.

Example.

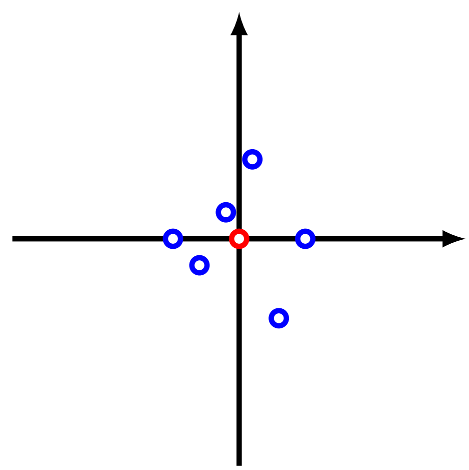
$$A = \begin{bmatrix} 5.0 & 6.0 \\ 7.0 & 2.0 \\ 4.0 & 4.0 \\ 3.0 & 5.0 \\ 6.0 & 8.0 \\ 8.0 & 5.0 \end{bmatrix}$$

$$\text{mean} = \begin{bmatrix} 5.5 & 5.0 \end{bmatrix}$$



$$\widetilde{A} = \begin{bmatrix} -0.5 & 1.0 \\ 1.5 & -3.0 \\ -1.5 & -1.0 \\ -2.5 & 0.0 \\ 0.5 & 3.0 \\ 2.5 & 0.0 \end{bmatrix}$$

$$\text{mean} = \begin{bmatrix} 0.0 & 0.0 \end{bmatrix}$$



Definition

The *covariance matrix* of a data matrix A is the matrix

$$C_A = \frac{1}{N}\,\widetilde{A}^T\widetilde{A}$$

Proposition

If $A = \begin{bmatrix} X_1 & \dots & X_M \end{bmatrix}$ is a data matrix then

$$C_A = \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \dots & \mathrm{Cov}(X_1, X_M) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \dots & \mathrm{Cov}(X_2, X_M) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_M, X_1) & \mathrm{Cov}(X_M, X_2) & \dots & \mathrm{Var}(X_M) \end{bmatrix}$$

Note:

If $C = \underbrace{\begin{bmatrix} c_1 & \dots & c_n \end{bmatrix}}_{\text{columns of } C}$ $\quad R = \begin{bmatrix} r_1 \\ \vdots \\ r_m \end{bmatrix} \Big\}$ rows of R

then

$$RC = \begin{bmatrix} r_1c_1 & r_1c_2 & \cdots & r_1c_n \\ r_2c_1 & r_2c_2 & \cdots & r_2c_n \\ \vdots & \vdots & & \vdots \\ \underbrace{r_mc_1}_{} & r_mc_2 & \cdots & r_mc_n \end{bmatrix}$$

multiplication of a row and a column

$$\widetilde{A} = \begin{bmatrix} \widetilde{X}_1 & \widetilde{X}_2 & \cdots & \widetilde{X}_M \end{bmatrix}$$

$$\widetilde{A}^T = \begin{bmatrix} \widetilde{X}_1^T \\ \widetilde{X}_2^T \\ \vdots \\ \widetilde{X}_M^T \end{bmatrix} \longrightarrow \frac{1}{N}\widetilde{A}^T\widetilde{A} = \frac{1}{N}\begin{bmatrix} \widetilde{X}_1^T\widetilde{X}_1 & \cdots & \widetilde{X}_1^T\widetilde{X}_M \\ \vdots & & \vdots \\ \widetilde{X}_M^T\widetilde{X}_1 & \cdots & \widetilde{X}_M^T\widetilde{X}_M \end{bmatrix} = \begin{bmatrix} \mathrm{Cov}(x_1, x_1) & \cdots & \mathrm{Cov}(x_1, x_M) \\ \vdots & & \vdots \\ \mathrm{Cov}(x_M, x_1) & \cdots & \mathrm{Cov}(x_M, x_M) \end{bmatrix}$$

Note. For any matrix A the matrix C_A is

- symmetric
- positive semidefinite

Total variance and trace

Definition

If $A = \begin{bmatrix} X_1 & \dots & X_M \end{bmatrix}$ is a data matrix then the *total variance* of A is the number

$$\mathrm{Var}(A) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_M)$$

Definition

For a square matrix

$$B = \begin{bmatrix} b_{11} & \dots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & \dots & b_{nn} \end{bmatrix}$$

the *trace* of B is the number

$$\operatorname{tr} B = b_{11} + b_{22} + \ldots + b_{nn}$$

Example:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \qquad \operatorname{tr} A = 1 + 5 + 9 = 15$$

Note. If A is a data matrix and C_A is its covariance matrix then

$$\mathrm{Var}(A) = \operatorname{tr} C_A$$

Proposition

If A, B are $n \times n$ matrices then

1) If A, B are $n \times n$ matrices then $\operatorname{tr} AB = \operatorname{tr} BA$.
2) If A, P, B are $n \times n$ matrices such that $A = PBP^{-1}$ then $\operatorname{tr} A = \operatorname{tr} B$.

Proof: Let $A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$, $B = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nn} \end{bmatrix}$

1) $$\operatorname{tr} AB \underset{\text{check}}{=} \sum_{ij} a_{ij} b_{ji} = \sum_{ij} b_{ji} a_{ij} = \operatorname{tr} BA$$

2) $$\operatorname{tr} A = \operatorname{tr}(PBP^{-1}) = \operatorname{tr}(P(BP^{-1})) \overset{\text{by 1)}}{=} \operatorname{tr}((BP^{-1})P)$$
$$= \operatorname{tr}(B(PP^{-1}))$$
$$= \operatorname{tr}(B \cdot I) = \operatorname{tr}(B)$$

Note: It is not true that $\operatorname{tr}(AB) = \operatorname{tr}(A) \cdot \operatorname{tr}(B)$.

For example :

$$A = B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad AB = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\operatorname{tr}(AB) = 2 \qquad \operatorname{tr}(A) \cdot \operatorname{tr}(B) = 2 \cdot 2 = 4$$

Corollary

If a matrix A is diagonalizable,

$$A = PDP^{-1}$$

for some invertible matrix P and a diagonal matrix

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

then $\operatorname{tr} A = \operatorname{tr} D = \lambda_1 + \lambda_2 + \dots + \lambda_n$.