PROBLEM #1.

Compute the covarient matrix for the dataset D.

$$D = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 4 \end{bmatrix} \right\}$$

The covarient matrix of a $\mathbb{R}^D$ dataset is given by the formula.

$$\mathrm{Var}[D] = \frac{1}{N} \sum (x_i - \mu)(x_i - \mu)^T$$

$$\mathrm{Var}\, X = \frac{1}{N} \sum^{N} (x_i - \bar{x})^2$$

$$\mathrm{Cov}\, X,Y = E[(x - \mu_x)(y - \mu_y)]$$

$$\mu_x = \frac{1}{N} \sum^{N} x_i$$

The 2D Covariance Matrix is,

$$\left[\begin{array}{c|c} \mathrm{Var}X & \mathrm{Cov}X,Y \\ \hline \mathrm{Cov}Y,X & \mathrm{Var}\,Y \end{array}\right]$$

Calculations,

$$\mu_x = \frac{1+5}{2} = 3$$

$$\mu_y = \frac{2+4}{2} = 3.$$

$$\mathrm{Var}X = \frac{(1-3)^2 + (5-3)^2}{2} = 4$$

$$\mathrm{Var}Y = \frac{(2-3)^2 + (4-3)^2}{2} = 1$$

$$\mathrm{Cov}\, X,Y = \frac{(1-3)(2-3) + (5-3)(4-3)}{2} = 2$$

$$\mathrm{Cov}\, Y.x = 2$$

$$\text{Covariance Matrix} = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$$

PROBLEM 2.

Given the covariance matrix of dataset D what is the covariance matrix if we multiply every vector by 2.

$$C = \frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})(X_i - \overline{X})^T \qquad C_D = \begin{bmatrix} 3 & 2 \\ 2 & 4 \end{bmatrix}$$

$$\begin{bmatrix} x_1 - \overline{x} \\ x_2 - \overline{x} \end{bmatrix} \quad \begin{bmatrix} x_1 - \overline{x}_1 & x_2 - \overline{x}_2 \end{bmatrix}$$

$$\begin{bmatrix} (x_1-\overline{x}_1)(x_1-\overline{x}_1) & (x_1-\overline{x})(x_2-\overline{x}) \\ (x_2-\overline{x}_1)(x_1-\overline{x}_2) & (x_2-\overline{x})(x_2-\overline{x}) \end{bmatrix} = \begin{bmatrix} \text{Var}X & \text{Cov}X,Y \\ \text{Cov}Y,X & \text{Var}Y \end{bmatrix}$$

$$\overline{x}_1 = \frac{1}{N}\sum^{N} x$$

$$\overline{x}_2 = \frac{1}{N}\sum^{N} x$$

When the dataset is scaled

$$\text{scaled } \overline{x}_1 = \frac{1}{N}\sum^{N} c \cdot x$$

$$= \frac{c}{N}\sum^{N} x.$$

same for $\overline{x}_2$.

Covariance matrix scaled by a constant,

$$C = \frac{1}{N}\sum_{i=1}^{N}(c \cdot X_i - c \cdot \overline{x}_1)(c \cdot X_i - c \cdot \overline{x}_2)$$

$$= \frac{1}{N}\sum_{i=1}^{N} c \cdot (X_i - \overline{x}_1) \cdot c \cdot (X_i - \overline{x}_2)$$

$$= \frac{c^2}{N}\sum_{i=1}^{N}(X_i - \overline{x})(x_i - \overline{x})^T$$

Thus scaling the Dataset D by a factor c means,

cov-matrix = c^2 cov-matrix

$@c=2$: $\quad c^2\begin{bmatrix}3 & 2\\ 2 & 4\end{bmatrix}$

$$2^2\begin{bmatrix}3 & 2\\ 2 & 4\end{bmatrix} = \begin{bmatrix}12 & 8\\ 8 & 16\end{bmatrix}$$

PROBLEM #3.

Given dataset D, D's covariance, calculate the covariance when adding $[2\ 2]^T$ to each element.

$$D = \left\{\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}7\\4\end{bmatrix}\right\} \qquad C_D = \begin{bmatrix}9 & 3\\ 3 & 1\end{bmatrix}$$

$$\bar{x}_1 = \frac{1}{N}\sum^N\left(x + \begin{bmatrix}2\\2\end{bmatrix}\right)$$

$$= \frac{1}{N}\left(\sum^N x + \sum^N [2\ 2]^T\right)$$

$$= \frac{1}{N}\sum^N x + \frac{N}{N}\begin{bmatrix}2\\2\end{bmatrix}$$

$$= \frac{1}{N}\sum^N x + [2\ 2]^T$$

$$C = \frac{1}{N}\sum^N (x_i - \bar{x})(x_i - \bar{x})^T$$

$$= \frac{1}{N}\sum^N \left(\left(x_i + \begin{bmatrix}2\\2\end{bmatrix}\right) - \bar{x}\right)\left(\left(x_i + \begin{bmatrix}2\\2\end{bmatrix}\right) - \bar{x}\right)^T$$

$$= \frac{1}{N}\sum^N \left(x_i + \begin{bmatrix}2\\2\end{bmatrix} - \begin{bmatrix}2\\2\end{bmatrix} + \bar{x}\right)\left(x_i + \begin{bmatrix}2\\2\end{bmatrix} - \begin{bmatrix}2\\2\end{bmatrix} + \bar{x}\right)^T,$$ substituting $\bar{x}$ for $\frac{1}{N}\sum^N x + \begin{bmatrix}2\\2\end{bmatrix}$

$$= C. \quad \square.$$

PROBLEM #4

Write a valid covariance 2x2 matrix

$C = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, $\quad \text{Var}\,X = 0 \quad \text{Cov}\,X,Y = \text{Cov}\,Y,X$

$\text{Var}\,Y = 0 \qquad\qquad = 0$

PROBLEM #5

Given the covariance matrix of dataset D which statements are true? Datasets consist of 2x1 vectors.

$$C_D = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$

- [x] x and y are positively correlated, i.e, when x increases then y increases on average and vice versa.

- [ ] x and y are negatively correlated, i.e. when x increases then y decreases on average and vice versa.

- [ ] x and y are uncorrelated, i.e, when x increases then y does not change on average (and vice versa).