1. Derive the mathematical relationship between cosine similarity and Euclidean distance when each data object has an *L2* length of 1.

   Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be two vectors where each vector has an *L2* length of 1. The relationship between the Euclidean distance and cosine similarity is derived as follows:

$$d(\boldsymbol{u},\boldsymbol{v}) = \sqrt{\sum_{k=1}^{n}(u_k - v_k)^2}$$
$$= \sqrt{\sum_{k=1}^{n}(u_k^2 - 2u_k v_k + v_k^2)}$$
$$= \sqrt{1 - 2\cos(\boldsymbol{u},\boldsymbol{v}) + 1}$$
$$= \sqrt{2(1 - \cos(\boldsymbol{u},\boldsymbol{v}))}$$

2. We consider the following data points: (2, 19), (9, 6), (7, 15), (5, 12).

   a) Calculate the covariance matrix of this set of data.

   We denote the first attribute as *x*, and the second attribute as *y*.

$$\bar{x} = \frac{2+9+7+5}{4} = 5.75, \quad \bar{y} = \frac{19+6+15+12}{4} = 13$$
$$\sigma_x^2 = \frac{1}{3}[(2-5.75)^2 + (9-5.75)^2 + (7-5.75)^2 + (5-5.75)^2] = 8.917$$
$$\sigma_y^2 = \frac{1}{3}[(19-13)^2 + (6-13)^2 + (15-13)^2 + (12-13)^2] = 30$$
$$\text{cov}(x,y) = \frac{1}{3}[(2-5.75)(19-13) + (9-5.75)(6-13) +$$
$$(7-5.75)(15-13) + (5-5.75)(12-13)] = -14$$

   The covariance matrix is given by $\begin{bmatrix} 8.917 & -14 \\ -14 & 30 \end{bmatrix}$

   b) Calculate the correlation coefficient between the two attributes.

   The correlation coefficient between the two attributes is

$$r_{xy} = \frac{-14}{\sqrt{8.917}\sqrt{30}} = -0.86$$