Fundamentals of Data Science

Semester B 20-21

Tutorial 3

1 Derive the mathematical relationship between cosine similarity and Euclidean distance when each data object has an L_2 length of 1.

2 We consider the following similarity measure between two vectors **x** and **y**.

$$S(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - \mathbf{x} \cdot \mathbf{y}}$$

Show that this measure corresponds to the Jaccard coefficient when **x** and **y** are binary vectors.

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

a. Calculate the covariance matrix of this set of data.
b. Calculate the correlation coefficient between the two attributes.