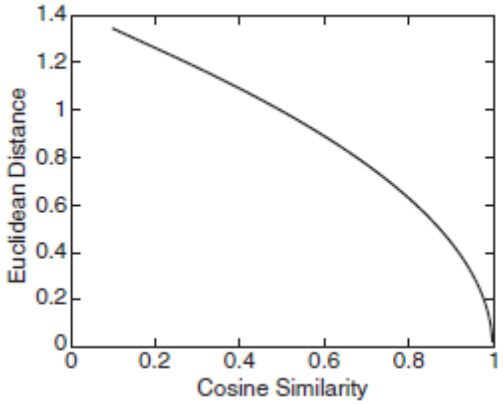
Tutorial 2

1. a. The following figure shows the relationship between the cosine similarity and the Euclidean distance for 100,000 randomly generated points that have been normalized to have an L_2 length of 1. What general observation can you make about the relationship between the two measures?



   b. 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 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.