Euclidean Distance:
linebreak Normalized Expectation and Variance in High Dimensions

Kanishk Navale (navalekanishk@gmail.com)

SEP. 08, 2024

1 Euclidean Distance

$$D(\mathbf{X}, \mathbf{Y}) = ||\mathbf{X} - \mathbf{Y}||_2 = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}, \tag{1}$$

where,

- $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^n$ are two vectors uniformly distributed over the range (a, b).
- $D \in \mathbb{R}$ is the Euclidean distance between $\mathbf{X}$ and $\mathbf{Y}$.

2 Normalized Euclidean Distance in High Dimensions

To make the Euclidean distance less sensitive to dimensionality, we normalize it using the expectation and variance in high-dimensional spaces.

2.1 Expected Euclidean Distance

For high dimensions, the expected Euclidean distance is approximately:

$$\mathbb{E}[D(\mathbf{X}, \mathbf{Y})] \approx (b - a) \cdot \sqrt{\frac{2n}{\pi}} \in \mathbb{R} \tag{2}$$

2.2 Variance of the Euclidean Distance

The variance of the Euclidean distance for high dimensions is given by:

$$\mathrm{Var}[D(\mathbf{X}, \mathbf{Y})] \approx \frac{(b - a)^2}{n} \cdot \left(1 - \frac{2}{\pi}\right) \in \mathbb{R} \tag{3}$$

The corresponding standard deviation is:

$$\sigma(D(\mathbf{X}, \mathbf{Y})) \approx \frac{(b - a)}{\sqrt{n}} \cdot \sqrt{1 - \frac{2}{\pi}} \in \mathbb{R} \tag{4}$$

2.3 Normalized Euclidean Distance

The normalized Euclidean distance, which is insensitive to the dimensionality, is:

$$D_{\mathrm{norm}}(\mathbf{X}, \mathbf{Y}) = \frac{D(\mathbf{X}, \mathbf{Y}) - (b - a)\sqrt{\frac{2n}{\pi}}}{\frac{(b-a)}{\sqrt{n}} \cdot \sqrt{1 - \frac{2}{\pi}}} \in \mathbb{R} \tag{5}$$

This distance is dimension-independent and accounts for both the expected value and variance of the Euclidean distance in high-dimensional spaces.