Euclidean Distance:
Expectation, Variance, and Standard Deviation

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 in range $(a, b), a << b$.
- $D \in \mathbb{R}$ is the Euclidean distance between $\mathbf{X}$ and $\mathbf{Y}$.

2 Derivation of Expectation

To compute the expectation of the Euclidean distance $D(\mathbf{X}, \mathbf{Y})$, we first calculate the expected value of the square of the distance.

2.1 Expected Squared Difference for One Dimension

For each component $x_i \in \mathbf{X}_{i=1,\dots,n}$ and $y_i \in \mathbf{Y}_{i=1,\dots,n}$, the expectation of the square of the difference uniformly distributed over (a, b) is:

$$\mathbb{E}[(x_i - y_i)^2] = \frac{(b-a)^2}{6} \in \mathbb{R} \tag{2}$$

2.2 Expected Squared Euclidean Distance

As the squared Euclidean distance is the sum of the squared differences over all n dimensions, its expectation is:

$$\mathbb{E}[D(\mathbf{X}, \mathbf{Y})^2] = n \cdot \frac{(b-a)^2}{6} \in \mathbb{R} \tag{3}$$

2.3 Expected Euclidean Distance

Taking the square root of the expected squared distance, the expectation of the Euclidean distance is as follows:

$$\mathbb{E}[D(\mathbf{X}, \mathbf{Y})] = \sqrt{\frac{n}{6}} \cdot (b-a) \in \mathbb{R} \tag{4}$$

3 Derivation of Variance

The variance of the Euclidean distance $D(\mathbf{X}, \mathbf{Y})$ is obtained by first calculating the variance of the squared Euclidean distance.

3.1 Variance of the Squared Difference

The variance of the squared difference $(x_i - y_i)^2$ is given by:

$$\mathrm{Var}[(x_i - y_i)^2] = \mathbb{E}[(x_i - y_i)^4] - \left(\mathbb{E}[(x_i - y_i)^2]\right)^2 \tag{5}$$

For uniformly distributed x_i and y_i over (a, b), we have:

$$\mathbb{E}[(x_i - y_i)^4] = \frac{(b-a)^4}{5} \in \mathbb{R} \tag{6}$$

Thus, the variance of $(x_i - y_i)^2$ is:

$$\mathrm{Var}[(x_i - y_i)^2] = (b-a)^4 \left(\frac{1}{5} - \frac{1}{36}\right) = (b-a)^4 \cdot \frac{31}{180} \in \mathbb{R} \tag{7}$$

3.2 Variance of the Squared Euclidean Distance

The squared Euclidean distance is the sum of the squared differences across all n dimensions, so its variance is:

$$\mathrm{Var}[D(\mathbf{X}, \mathbf{Y})^2] = n \cdot (b-a)^4 \cdot \frac{31}{180} \in \mathbb{R} \tag{8}$$

4 Derivation of Standard Deviation

Finally, the variance of the Euclidean distance $D(\mathbf{x}, \mathbf{y})$ can be approximated using the delta method:

$$\mathrm{Var}[D(\mathbf{X}, \mathbf{Y})] \approx \frac{1}{4} \cdot \frac{\mathrm{Var}[D(\mathbf{X}, \mathbf{Y})^2]}{\mathbb{E}[D(\mathbf{X}, \mathbf{Y})^2]} \tag{9}$$

We know that:

$$\mathbb{E}[D(\mathbf{X}, \mathbf{Y})^2] = n \cdot \frac{(b-a)^2}{6} \in \mathbb{R} \tag{10}$$

Substituting, we get:

$$\mathrm{Var}[D(\mathbf{X}, \mathbf{Y})] = \frac{31}{120} \cdot (b-a)^2 \in \mathbb{R} \tag{11}$$

Therefore, the standard deviation is:

$$\sigma(n) = \sqrt{\frac{31}{120}} \cdot (b-a) \approx 0.507 \cdot (b-a) \in \mathbb{R} \tag{12}$$