Bounds on the Euclidean distance between two probability mass functions

vishalcr • 29 May 2025

We have two vectors $\underline{x} \in \mathbb{R}^n$ and $\underline{y} \in \mathbb{R}^n$ that represent probability mass functions (pmf's). Any pmf $\underline{p} \in \mathbb{R}^n$ must satisfy the following two properties:

$$\begin{aligned} \sum_{i=1}^{n} p_i &= 1 \\ p_i &\geq 0,\ \forall i \end{aligned} \tag{1}$$

where p_i represents the probability that the discrete random variable x takes the value i, i.e., $Pr\{x = i\} = p_i$ and $i \in \{1, \ldots, n\}$.

The Euclidean distance (d) between the two pmf's $\underline{x}$ and $\underline{y}$ can be expressed as

$$d^2 = \sum_{i=1}^{n} (x_i - y_i)^2 \tag{2}$$

This can be expanded to

$$\begin{aligned} d^2 &= \sum_{i=1}^{n} \left(x_i^2 + y_i^2 - 2x_i y_i\right) \\ &= \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} y_i^2 - 2\sum_{i=1}^{n} x_i y_i \end{aligned} \tag{3}$$

Because of the definition in (1)

$$\begin{aligned} d^2 &\leq 1 + 1 - 2\sum_{i=1}^{n} x_i y_i \\ &\leq 2 - 2\left(\sum_{i=1}^{n} x_i y_i\right) \end{aligned} \tag{4}$$

In (4), as due to the second equation in (1), $\sum_{i=1}^{n} x_i y_i \geq 0$, therefore

$$\begin{aligned} d^2 &\leq 2 \\ d &\leq \sqrt{2} \end{aligned} \tag{5}$$

Hence, as $d \geq 0$ by definition as a distance metric, *the Euclidean distance between two pmf vectors* d *is always bounded within*

$$0 \leq d \leq \sqrt{2} \tag{6}$$