Wasserstein Distance to Uniform Distribution (WDUD)

Given $\{v_n\}_{n=1}^N$ as properties for N molecules. (In the paper, these are called y_n.)

Define:

$$v_{\min} = \min_{1 \le n \le N} v_n$$
$$v_{\max} = \max_{1 \le n \le N} v_n$$

A uniform distribution is defined as the function

$$p_U(v) = \frac{1}{\int_{v_{\min}}^{v_{\max}} dv}$$

The cumulative distribution function is

$$P(v) = \int_{v_{\min}}^{v} p(v')\, dv'$$

so for a uniform distribution, the cumulative distribution is merely

$$P_U(v) = \frac{v - v_{\min}}{v_{\max} - v_{\min}}$$

The distribution of the properties is estimated by assuming that the probability of each observation is equal, $\frac{1}{N}$. Sort the property values, obtaining a new property distribution

$$\{v'_n\} = \text{sort}\{v_n\}$$

Then

$$v'_1 = v_{\min}$$
$$v'_N = v_{\max}$$

$$P_V(v) = \begin{cases} 0 & v < v_{\text{min}} \\ \frac{1}{N} & v'_1 \leq v < v'_2 \\ \vdots & \\ \frac{k-1}{N} & v'_{k-1} \leq v < v'_k \\ \vdots & \\ \frac{N-1}{N} & v'_{N-1} \leq v < v'_N \\ 1 & v > v_{\text{max}} \end{cases}$$

So the formula you need is simply expressible as

$$P_V(v) = \begin{cases} 0 & v < v'_1 \\ \frac{k-1}{N} & v'_{k-1} \leq v < v'_k \\ 1 & v'_N \leq v \end{cases}$$

You can evaluate the integral numerically to get the WDUD,

$$\text{WDUD} = \int_{v_{\text{min}}}^{v_{\text{max}}} \left| P_V(v) - P_U(v) \right| dv$$

Now when there are multiple properties, you first define each property in a normalized way. So we define

$$\{v''_n\} = \left\{ \frac{v'_n - v_{\text{min}}}{v_{\text{max}} - v_{\text{min}}} \right\}$$

and then compute the WDUD of $\{v''_n\}$. Compute the WDUD of all the properties then average (take their mean) to get the multi-property WDUD.

If the data is normalized, then

$$P_U(v) = \begin{cases} 0 & v < 0 \\ v & 0 \leq v \leq 1 \\ 1 & 1 < v \end{cases}$$