

1(a)

Earth Mover Distance

Sequenz $p = \{(p_i, w_{p_i}) \mid p_i \in \mathbb{R}^d \text{ und } w_{p_i} \in \mathbb{R}^{>0}\}$

Grunddistanz $d = \text{euclid}(x, y)$

$p = \{(p_i, w_{p_i}), \dots\}$ Erdhügel $p_i = \text{Position}$, $w_{p_i} = \text{Volumen Hügel}$

$q = \{(q_i, w_{q_i}), \dots\}$ Erdlöcher $q_i = \text{Position}$, $w_{q_i} = \text{Volumen Loch}$

$F = [f_{ij}]$ konkreter Transport, f_{ij} Menge von $p_i \xrightarrow{w_{p_i}} q_j$

$$\text{Kosten}(p, q, F) = \sum_{i=1}^m \sum_{j=1}^n d(p_i, q_j) \cdot f_{ij}$$

$$f_{ij} > 0, \quad \sum_{j=1}^n f_{ij} \leq w_{p_i}, \quad \sum_{i=1}^m f_{ij} \leq w_{q_j}$$

$$\text{Ziel: } \sum_{i=1}^m \sum_{j=1}^n f_{ij} = \min \left(\sum_{i=1}^m w_{p_i}, \sum_{j=1}^n w_{q_j} \right), \quad \text{Es darf etwas in dem Hügel oder Lochern überbleiben.}$$

$$d_{\text{EMD}} = \frac{\min_{[f_{ij}]} \left(\sum_{i=1}^m \sum_{j=1}^n d(p_i, q_j) \cdot f_{ij} \right)}{\sum_{i=1}^m \sum_{j=1}^n f_{ij}}$$

(1) Normieren des Histogramms 1 & 2

$$B_{H_1}(1, 0.2), B_{H_1}(2, 0.6), B_{H_1}(3, 0.1), B_{H_1}(4, 0.1)$$

$$B_{H_2}(1, 0.25), B_{H_2}(2, 0.05), B_{H_2}(3, 0.65), B_{H_2}(4, 0.05)$$

$$\text{Es gilt für } H_1 \text{ und } H_2: \sum_i w_i = 1$$

(2) Sequenzierung (Positionen $\binom{x}{y}$) (Positionen ergeben sich nach Aufgabendstellung aus der Beschriftung)

$$H_1 = \left\{ \binom{1}{0}, 0.2, \binom{2}{0}, 0.6, \binom{3}{0}, 0.1, \binom{4}{0}, 0.1 \right\}$$

$$H_2 = \left\{ \binom{1}{0}, 0.25, \binom{2}{0}, 0.05, \binom{3}{0}, 0.65, \binom{4}{0}, 0.05 \right\}$$

(3) Move

$d_{1,2}$	H_{11}	H_{12}	H_{13}	H_{14}
H_{21}	0	1	2	3
H_{22}	1	0	1	2
H_{23}	2	1	0	1
H_{24}	3	2	1	0

I1) Schiebe alles in den nächsten Nachbarn, da Kosten=0.

f_{ij}	H_{21}	H_{22}	H_{23}	H_{24}
H_{21}	0.2	0	0	0
H_{12}	0	0.05	0	0
H_{23}	0	0	0.1	0
H_{14}	0	0	0	0.05

$$H'_1 = \left\{ \left(\binom{1}{0}, 0 \right), \left(\binom{2}{0}, 0.55 \right), \left(\binom{3}{0}, 0 \right), \left(\binom{4}{0}, 0.05 \right) \right\}$$

$$H'_2 = \left\{ \left(\binom{1}{0}, 0.05 \right), \left(\binom{2}{0}, 0 \right), \left(\binom{3}{0}, 0.55 \right), \left(\binom{4}{0}, 0 \right) \right\}$$

I

I2)

I: Kosten $3 \cdot 0.05 + 1 \cdot 0.55 = 0.65$

II: Kosten $1 \cdot 0.05 + 1 \cdot 0.05 + 1 \cdot 0.50 = 0.6$

=> Da bereits die nächsten Nachbarn befüllt sind könnte wieder das Selbe gemacht werden (I).

Aber dadurch müsste W_{p_4} über 3 Bins zu W_{q_4} getragen werden. Anders könnte von $W_{p_2} \xrightarrow{0.05} W_{q_1}$ und $W_{p_4} \xrightarrow{0.05} W_{q_3}$ sowie $W_{p_2} \xrightarrow{0.5} W_{q_3}$.

Dadurch würden die Wege nimmert, da jeder Haufen nur über einen Bin getragen werden muss.

Auch werden so 3 Schritte anstelle von 4 (II) gemacht.

f_{ij}	H_{21}	H_{22}	H_{23}	H_{24}	
H_{21}	0.2	0	0	0	0.2
H_{12}	0.05	0.05	0.5	0	0.6
H_{23}	0	0	0.1	0	0.1
H_{14}	0	0	0.05	0.05	0.1
	0.25	0.05	0.65	0.05	$\Sigma=1$

Kosten(H_1, H_2, F) =

$$0 \cdot 0.2 + 1 \cdot 0.05 + 0 \cdot 0.05 + 1 \cdot 0.5 + 0 \cdot 0.1 + 1 \cdot 0.05 + 0 \cdot 0.05 = 0.6$$

1(6) Da die Distanz ist minimal und mein Verfahren würde Histogramme mit großer Anzahl die minimale Distanz angehen.

Das Vorgehen dabei ist, immer die nächsten Nachbarn mit einer kleinen Menge zu füllen bis diese voll sind. Große Mengen sollen bis zum Ende aufbewahrt werden. Da kleine Mengen durch die nächsten Nachbarn eine kleine Gewichtung bekommen wird lediglich im letzten Schritt eine große Menge "stark" gewichtet. Daher werden die Kosten gering und minimal. Dieser Zusammenhang kommt daher, dass die Distanzen maßgeblich die Kosten mitbestimmen.

Sie sind stets größer als die normierten w_{q_i} , w_{p_i} , daher fallen sie für größere stärker ins Gewicht als für kleine Mengen.