練習

1. 令 p1 =(2, 3), p2 = (5, 1), p3=(1, 4), p4=(0, 1)，請分別算出

   甲、歐幾里德距離矩陣

   乙、閔可夫斯基之 L1, 與 L∞之距離矩陣

Solution:

Eq: Euclidean distance

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{k=1}^{n} (x_k - y_k)^2}$$

$$\mathrm{d(p1, p2)} = \sqrt{(2-5)^2 + (3-1)^2} = \sqrt{9+4} = 3.606$$
$$\mathrm{d(p1, p3)} = \sqrt{(2-1)^2 + (3-4)^2} = \sqrt{1+1} = 1.141$$
$$\mathrm{d(p1, p4)} = \sqrt{(2-0)^2 + (3-1)^2} = \sqrt{4+4} = 2.828$$
$$\mathrm{d(p2, p3)} = \sqrt{(5-1)^2 + (1-4)^2} = \sqrt{16+9} = 5$$
$$\mathrm{d(p2, p4)} = \sqrt{(5-0)^2 + (1-1)^2} = \sqrt{25+0} = 5$$
$$\mathrm{d(p3, p4)} = \sqrt{(1-0)^2 + (4-1)^2} = \sqrt{1+9} = 3.162$$

d(pi, pj)	P1	P2	P3	P4
P1	0	3.606	1.141	2.828
P2	3.606	0	5	5
P3	1.141	5	0	3.162
P4	2.828	5	3.162	0

Eq: Minkowski distance

$$d(\mathbf{x}, \mathbf{y}) = \left(\sum_{k=1}^{n} |x_k - y_k|^r \right)^{1/r}$$

L1: 公式變成 r = 1

$$d(x, y) = \left(\sum_{k=1}^{n} |x_k - y_k|^1 \right)^{1/1}$$

$$\mathrm{d(p1, p2)} = |2-5| + |3-1| = 3+2 = 5$$
$$\mathrm{d(p1, p3)} = |2-1| + |3-4| = 1+2 = 3$$
$$\mathrm{d(p1, p4)} = |2-0| + |3-1| = 2+2 = 4$$
$$\mathrm{d(p2, p3)} = |5-1| + |1-4| = 4+3 = 7$$
$$\mathrm{d(p2, p4)} = |5-0| + |1-1| = 5+0 = 5$$
$$\mathrm{d(p3, p4)} = |1-0| + |4-1| = 1+3 = 4$$

d(pi, pj)	P1	P2	P3	P4

P1	0	5	3	4
P2	5	0	7	5
P3	3	7	0	4
P4	4	5	4	0

L∞: 公式變成

$$\lim_{x \to \infty}\left(\sum_{k=1}^{n}|x_k - y_k|^r\right)^{1/r} = \max_{i=1}^{n}|x_k - y_k|$$

$$\mathrm{d}(\mathrm{p1}, \mathrm{p2}) = \max\{|2-5|, |3-1|\} = \max\{3, 2\} = 3$$
$$\mathrm{d}(\mathrm{p1}, \mathrm{p3}) = \max\{|2-1|, |3-4|\} = \max\{1,1\} = 1$$
$$\mathrm{d}(\mathrm{p1}, \mathrm{p4}) = \max\{|2-0|, |3-1|\} = \max\{2, 2\} = 2$$
$$\mathrm{d}(\mathrm{p2}, \mathrm{p3}) = \max\{|5-1|, |1-4|\} = \max\{4, 3\} = 4$$
$$\mathrm{d}(\mathrm{p2}, \mathrm{p4}) = \max\{|5-0|, |1-1|\} = \max\{5,0\} = 5$$
$$\mathrm{d}(\mathrm{p3}, \mathrm{p4}) = \max\{|1-0|, |4-1|\} = \max\{1,3\} = 3$$

d(pi, pj)	P1	P2	P3	P4
P1	0	3	1	2
P2	3	0	4	5
P3	1	4	0	3
P4	2	5	3	0

2. 令 x = {0, 1, 1, 0, 0, 0, 1, 0, 1, 1}, y = {1, 0, 1, 0, 1, 0, 0, 1, 0, 1}，請算出
   甲、SMC
   乙、Jaccard

Solution

$$SMC = \frac{\text{同時為 1 或同時為 0 的屬性個數}}{\text{屬性個數}} = \frac{f_{11} + f_{00}}{f_{01} + f_{10} + f_{11} + f_{00}}$$

是計算所有向量中同時為 1 或同時為 0 的元素個數比例。

f00 = 2, f01 = 3, f10 = 3, f11 = 2

SMC = (f11+f00)/ (f01+f10+f11+f00) = (2 + 2) / (2 + 3 + 3 + 2) = 0.4

$$J = \frac{\text{同時為 1 的屬性個數}}{\text{不同時為 0 的屬性個數}} = \frac{f_{11}}{f_{01} + f_{10} + f_{11}}$$

J = f11 / (f01+f10+f11) = 2 / 8 = 0.25

3. 令 x = {3, 2, 1, 0, 1, 0, 1, 0, 1, 1},
4. y = {5, 3, 1, 0, 1, 2, 0, 4, 0, 1}

甲、cos(x, y)

乙、EJ

Solution

$$\cos(\mathbf{x},\mathbf{y}) = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|}$$

x • y = 3×5 + 2×3 + 1×1+ 0×0 + 1×1 + 0×2 + 1×0 + 0×4 + 1×0 + 1×1 =24

||x|| =

$\sqrt{3\times 3+2\times 2+1\times 1+0\times 0+1\times 1+0\times 0+1\times 1+0\times 0+1\times 1+1\times 1} = \sqrt{18} = 4.243$

||y|| =

$\sqrt{5\times 5+3\times 3+1\times 1+0\times 0+1\times 1+2\times 2+0\times 0+4\times 4+0\times 0+1\times 1} = \sqrt{57} = 7.550$

$$\cos(\mathrm{x},\mathrm{y}) = \frac{24}{4.243\times 7.550} = 0.749$$

$$EJ(\mathbf{x},\mathbf{y}) = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|^2+\|\mathbf{y}\|^2-\mathbf{x}\cdot\mathbf{y}}$$

$$\mathrm{EJ}(\mathrm{x},\mathrm{y}) = \frac{24}{4.243^2\times 7.550^2-24} = 0.471$$