

Exercises for Chapter 2

2.1 Given a numerical data vector, for which scales can you apply the following operations?

- a) bitwise exclusive or nominal, ordinal, interval, ratio
- b) subtraction interval, ratio
- c) subtraction followed by test for zero subtraction; interval, ratio
- d) sorting in descending order ordinal, interval, ratio
- e) computing a histogram ordinal, interval, ratio
- f) computing the discrete Fourier cosine transform interval, ratio

2.2 For two-dimensional data, find the unit circle, i.e. all points with a distance of one from the origin, for

- a) Euclidean distance $x^2 + y^2 = 1 \Rightarrow$ circle with radius 1
- b) city block distance $|x| + |y| = 1 \Rightarrow$ rhombus with the edge points $(0, \pm 1)$ and $(\pm 1, 0)$
- c) Hamming distance $(x=0 \wedge y \neq 0) \vee (x \neq 0 \wedge y=0) \Rightarrow$ coordinate axes without origin
- d) supremum norm $\max\{|x|, |y|\} = 1 \Rightarrow$ square with the edge points $(\pm 1, \pm 1)$ and $(\pm 1, 0)$
- e) matrix norm with $A = \begin{pmatrix} 0.1 & 0 \\ 0 & 1 \end{pmatrix}$ $0.1x^2 + y^2 = 1 \Rightarrow$ horizontal ellipse through the points $(0, \pm 1)$ and $(\pm \sqrt{10}, 0)$
- f) matrix norm with $A = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}$ $\Rightarrow (1-2)(x-y)^2 = 1 \Rightarrow y = x \pm 1 \Rightarrow$ two diagonals above and below the main diagonal, at a distance of $\sqrt{2}/2$

2.3 Consider sequences of the length three using the symbols \oplus and \ominus , for example $\oplus \ominus \ominus$.

- a) Find a pair of such sequences whose Hamming distance is equal to their edit distance. $\{\oplus\oplus\ominus, \ominus\ominus\oplus\}$: Hamming distance = 3 = edit distance
- b) Find a pair of such sequences whose Hamming distance is different from their edit distance. $\{\oplus\ominus\oplus, \ominus\oplus\ominus\}$: Hamming = 3, edit = 2 (remove first, append last)
- c) Find all possible sets of arbitrary number of such sequences whose Hamming distance matrices are equal to their edit distance matrices. All sets of sequences that do not contain both $\ominus\ominus\ominus$ and $\oplus\oplus\oplus$

2.4 Consider cosine similarity for two-dimensional non-negative feature data.

- a) Find all points that have a cosine similarity of 1 with $(1, 1)$.
- b) Find all points that have a cosine similarity of $0.5\sqrt{2}$ with $(1, 1)$.
- c) Find all points that have a cosine similarity of $0.3\sqrt{10}$ with $(1, 1)$.
- d) How do you interpret these results?

Hamming distance

It measures the minimum number of substitutions required to change one string into the other.

Ex.:

"Karolin" and "Kathrin" → distance 3

Edit distance

How dissimilar 2 strings are to one another / measured by counting the minimum number of operations required to transform one string into the other.

REMOVE, APPEND, EDIT

a) Find all points that have a cosine similarity of 1 with $(1, 1)$.

$$\frac{x \cdot 1 + y \cdot 1}{\sqrt{(x^2+y^2)(1^2+1^2)}} = \frac{x+y}{\sqrt{2(x^2+y^2)}} = 1 \Rightarrow$$

$$x^2 + 2xy + y^2 = 2x^2 + 2y^2 \Rightarrow x^2 - 2xy + y^2 = 0 \Rightarrow$$

$$(x-y)^2 = 0 \Rightarrow x=y \text{ except } x=y=0$$

Main diagonal

b) Find all points that have a cosine similarity of $0.5\sqrt{2}$ with $(1, 1)$.

$$\frac{x+y}{\sqrt{2(x^2+y^2)}} = \frac{1}{\sqrt{2}} \Rightarrow \sqrt{2}(x+y) = \sqrt{2(x^2+y^2)} \Rightarrow$$

$$x^2 + 2xy + y^2 = x^2 + y^2 \Rightarrow 2xy = 0 \Rightarrow x = 0 \text{ or}$$

$$y=0 \text{ except } x=y=0$$

c) Find all points that have a cosine similarity of $0.3\sqrt{10}$ with $(1, 1)$.

$$\frac{x+y}{\sqrt{2(x^2+y^2)}} = \sqrt{\frac{9}{10}} \Rightarrow x^2 + 2xy + y^2 = \frac{9}{5}x^2 + \frac{9}{5}y^2 \Rightarrow$$

$$\frac{4}{5}x^2 - 2xy + \frac{4}{5}y^2 = 0 \Rightarrow x^2 - \frac{5}{2}xy + y^2 = 0 \Rightarrow$$

$$x = \frac{5}{4}y \pm \sqrt{\frac{25}{16}y^2 - y^2} = \frac{5}{4}y \pm \frac{3}{4}y = 2y \text{ or } \frac{1}{2}y \text{ except } x=y=0$$



d) How do you interpret these results?

(positive parts of) two lines through the origin with angles α from the main diagonal, where α increases with decreasing similarity:
 0° (minimum) for similarity 1, 22.5° (measured) for similarity $0.3\sqrt{10}$, 45° (maximum) for similarity $\frac{1}{2}\sqrt{2}$