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

   we need subtraction, so interval, ratio

d) sorting in descending order

   we need comparisons (<,>), so ordinal, interval, ratio

e) computing a histogram

   we need to compare with bin edges (<,>), so ordinal, interval, ratio

f) computing the discrete Fourier cosine transform

   we need multiplication, but only by constants, so 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 $(\pm 1, \mp 1)$

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 $A = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}$

   $\begin{pmatrix} x & y \end{pmatrix} \cdot \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = (x - y)^2 = 1 \Rightarrow y = x \pm 1 \Rightarrow$ 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.

   for example $\{\oplus\oplus\oplus, \ominus\ominus\ominus\}$: Hamming distance 3, edit distance 3

b) Find a pair of such sequences whose Hamming distance is different from their edit distance.

   $\{\oplus\ominus\oplus, \ominus\oplus\ominus\}$: Hamming distance 3, edit distance 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.

   The only pair of sequences whose Hamming distance is different from their edit distance is the one given in (b). So, the answer for (c) is: all sets of sequences that do not contain both $\oplus\ominus\oplus$ and $\ominus\oplus\ominus$.

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

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

   $\frac{x\cdot 1+y\cdot 1}{\sqrt{(x^2+y^2)(1^2+1^2)}} = \frac{x+y}{\sqrt{2(x^2+y^2)}} = 1 \quad\Rightarrow\quad x^2+2xy+y^2 = 2x^2+2y^2 \quad\Rightarrow$ $x^2-2xy+y^2=0 \quad\Rightarrow\quad (x-y)^2=0 \quad\Rightarrow\quad x=y$ except $x=y=0$

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

   $\frac{x+y}{\sqrt{2(x^2+y^2)}} = \frac{1}{\sqrt{2}} \quad\Rightarrow\quad x^2+2xy+y^2 = x^2+y^2 \quad\Rightarrow\quad 2xy=0 \quad\Rightarrow$ $x=0$ or $y=0$ except $x=y=0$

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

   $\frac{x+y}{\sqrt{2(x^2+y^2)}} = \sqrt{\frac{9}{10}} \quad\Rightarrow\quad x^2+2xy+y^2 = \frac{9}{5}x^2+\frac{9}{5}y^2 \quad\Rightarrow\quad \frac{4}{5}x^2-2xy+$ $\frac{4}{5}y^2=0 \quad\Rightarrow\quad x^2-\frac{5}{2}xy+y^2=0 \quad\Rightarrow\quad x=\frac{5}{4}y\pm\sqrt{\frac{25}{16}y^2-y^2} =$ $\frac{5}{4}y\pm\frac{3}{4}y = 2y$ or $\frac{1}{2}y$ 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, $22.5°$ (mean) for similarity $0.3\sqrt{10}$, $45°$ (maximum) for similarity $\frac{1}{2}\sqrt{2}$