

## Exercise 1

We can define the *edit distance*  $d_{\text{edit}}(w, w') : \Sigma^2 \rightarrow \mathbb{R}$  as follows. (Let  $w = w_1 \dots w_n$  and  $w' = w'_1 \dots w'_m$ )

$$d_{\text{edit}}(w, w') \mapsto \begin{cases} |w| & \text{if } |w'| = 0 \\ |w'| & \text{if } |w| = 0 \\ d_{\text{edit}}(w_2 \dots w_n, w'_2 \dots w'_m) & \text{if } w_1 = w'_1 \\ 1 + \min \begin{cases} d_{\text{edit}}(w_2 \dots w_n, w') \\ d_{\text{edit}}(w, w'_2 \dots w'_m) \\ d_{\text{edit}}(w_2 \dots w_n, w'_2 \dots w'_m) \end{cases} & \text{otherwise} \end{cases}$$

As this definition of  $d_{\text{edit}}$  works by removing at most the first character of each word, we can proof by induction the length of  $x, y, z \in \Sigma$ , that  $d_{\text{edit}}$  is a metric on  $\Sigma$ :

- Let  $|x| = |y| = |z| = 0$ . Therefore, also  $x = y = z$ .  
Then  $0 \leq d_{\text{edit}} = 0$ . Thus, Nonnegativity is given.  
Since  $x = y$ , also  $d_{\text{edit}}(x, y) = d_{\text{edit}}(y, x)$ . Thus, Symmetry is given.  
Since  $x = y = z$ , the Triangle Inequality  $d_{\text{edit}}(x, z) \leq d_{\text{edit}}(x, y) + d_{\text{edit}}(y, z) \Leftrightarrow 0 \leq 0 + 0$  is given.
- Let  $x = x_1 \dots x_n$ ,  $y = y_1 \dots y_m$ , and  $z = z_1 \dots z_o$ ,  $n, m, o \geq 1$ . For  $x' = x_2 \dots x_n$ ,  $y' = y_2 \dots y_m$ , and  $z' = z_2 \dots z_o$  Nonnegativity, Symmetry, and the Triangle Inequality of  $d_{\text{edit}}$  is given.
- Since  $n, m \geq 1$ , the second rule of Nonnegativity, namely  $d_{\text{edit}}(x, y) \Leftrightarrow x = y$  does not apply here.  
Since all  $d_{\text{edit}}(x', y')$ ,  $d_{\text{edit}}(x', y)$ ,  $d_{\text{edit}}(x, y')$  are non-negative, by definition of  $d_{\text{edit}}$ ,  $d_{\text{edit}}(x, y)$  must be non-negative as well. Therefore, the Nonnegativity of  $d_{\text{edit}}$  is proven.
- Symmetry
- Triangle Inequality

## Exercise 2

(a),(b)

(c),(d)

Coordinate	Label	(1,-2,0)	(4,-0.5,2)	(1,1.5,-2.5)	(-2,-1,-2)	(-4,-1,-1)
(-4,-2.1,-1)	-1	6.1	12.6	10.1	4.1	1.1
(-3.6,-1.4,0.2)	1	5.3999	10.3	10.2	4.2	1.9999
(1,-0.2,-0.3)	1	2.1	5.6	3.9	5.5	6.5
(0.3,-0.5,-0.5)	1	2.7	6.2	4.7	4.3	5.3
(-2, -3.5, -1)	-1	5.5	12.0	9.5	3.5	4.5
(-4.2, -4, 0.2)	1	7.4	13.5	13.3999	7.4	4.4
(-1.3, -0.1, -3)	1	7.1999	10.7	4.4	2.6	5.6
(-0.7, 0.9, -0.7)	1	5.3	8.8	4.1	4.5	5.4999
(1, 2, 1.4)	1	5.4	6.1	4.4	9.4	10.4
(2.6, -1.5, 0.2)	1	2.3	4.2	7.3	7.3	8.2999
(2, 4.3, -0.7)	-1	8.0	9.5	5.6	10.6	11.6
(0.6, 0.4, 0.2)	-1	3.0	6.1	4.2	6.2	7.2
(2.9, -1.7, 3.6)	-1	5.8	3.9	11.2	11.2	12.2
(3.6, 0.4, -2.5)	-1	7.5	5.8	3.7	7.5	10.5
(1.2, 4, 1.2)	-1	7.4	8.1	6.4	11.3999	12.3999
(-1, 0.5, 0.5)	-1	5.0	7.5	6.0	5.0	6.0
(3, 2.7, 2.3)	-1	9.0	4.5	8.0	13.0	14.0
(4, -3, 2.2)	-1	6.2	2.7	12.2	12.2	13.2
(0.1, 0.1, 3.5)	-1	6.5	6.0	8.3	8.7	9.7
(2.8, 1.2, 2.4)	-1	7.4	3.3	7.0	11.4	12.4
Classification (k=2)		1	-1	0	0	0
Classification (k=3)		1	-1	1	1	1

Table 1: Manhattan distance table for the 5 query points. Classifications for k=2 (a) and k=3 (b) are stated below.

Coordinate	Label	(1,-2,0)	(4,-0.5,2)	(1,1.5,-2.5)	(-2,-1,-2)	(-4,-1,-1)
(-4, -2.1, -1)	-1	5.1	8.692	6.341	2.491	1.1
(-3.6, -1.4, 0.2)	1	4.643	7.862	6.071	2.749	1.326
(1, -0.2, -0.3)	1	1.825	3.792	2.780	3.539	5.111
(0.3, -0.5, -0.5)	1	1.729	4.465	2.913	2.791	4.357
(-2, -3.5, -1)	-1	3.5	7.348	6.020	2.692	3.201
(-4.2, -4, 0.2)	1	5.575	9.095	8.036	4.322	3.237
(-1.3, -0.1, -3)	1	4.231	7.297	2.846	1.516	3.478
(-0.7, 0.9, -0.7)	1	3.434	5.598	2.547	2.643	3.819
(1, 2, 1.4)	1	4.238	3.950	3.931	5.436	6.305
(2.6, -1.5, 0.2)	1	1.688	2.489	4.341	5.123	6.726
(2, 4.3, -0.7)	-1	6.417	5.859	3.475	6.766	8.011
(0.6, 0.4, 0.2)	-1	2.441	3.950	2.942	3.682	4.955
(2.9, -1.7, 3.6)	-1	4.082	2.282	7.145	7.473	8.322
(3.6, 0.4, -2.5)	-1	4.332	4.606	2.823	5.793	7.872
(1.2, 4, 1.2)	-1	6.122	5.360	4.469	6.743	7.541
(-1, 0.5, 0.5)	-1	3.240	5.315	3.741	3.082	3.674
(3, 2.7, 2.3)	-1	5.602	3.366	5.336	7.561	8.577
(4, -3, 2.2)	-1	3.852	2.507	7.165	7.592	8.845
(0.1, 0.1, 3.5)	-1	4.180	4.221	6.226	5.989	6.186
(2.8, 1.2, 2.4)	-1	4.386	2.118	5.228	6.873	7.914
Classification (k=2)		1	-1	1	0	0
Classification (k=3)		1	-1	1	1	-1

Table 2: Euclidian distance table for the 5 query points. Classifications for k=2 (a) and k=3 (b) are stated below.