1) Hammering dist

$$d(x,y) = |x_1 - y_1| + |x_2 - y_2| + \dots + |x_n - y_n|$$

$$= |0-0| + |1-1| + |0| + |1| + |0| + |0| + |0|$$
$$+ |0| + |0| + |1|$$

$$= 3$$

Jaccard similarity

$$J = \frac{f_{11}}{f_{01} + f_{10} + f_{11}}$$

$$f_{01} = 1, \quad f_{10} = 2, \quad f_{11} = 2, \quad f_{00} = 5$$

$$J = \frac{2}{1+2+2} = 0.4$$

2) a) $x = (1,1,1)$ $y = (2,2,2,2)$

Cosine

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

$x \cdot y = 1*2 + 1*2 + 1*2 + 1*2 = 8 \;\backslash\backslash$

$\|x\| = \sqrt{|x| + |x| + |x| + |x|} = 2 \backslash\backslash$

$\|y\| = \sqrt{2*2 + 2*2 + 2*2 + 2*2}$

$\quad = 4 \backslash\backslash$

$\cos(x,y) = \frac{8}{2 \times 4} = 1 \backslash\backslash$

Corelation

$$\cos(x,y) = \frac{\sum[(x_i - \bar{x}) \times (y_i - \bar{y})]}{\sqrt{\sum[(x_i - x)]^2} \times \sqrt{\sum[y_i - y)]^2}}$$

x	y	$x_i - x$	$y_i - y$	$(x_i - x)(y_i - y)$	$(x_i - x)^2$	$(y_i - y)^2$
1	2	0	0	0	0	0
1	2	0	0	0	0	0
1	2	0	0	0	0	0
1	2	0	0	0	0	0

$\cos(x,y) = 0 \backslash\backslash$

b) $x = (0,1,0,1)$ $y = (2,2,2,2)$

cosine

$x \cdot y = 0$

$\|x\| = \sqrt{2}$ $\|y\| = \sqrt{2}$

$\cos(x,y) = \dfrac{0}{2} = 0\!\!/\!\!/$

corelation

$\bar{x} = \dfrac{2}{4} = 0\!\!/\!\!/$ $\bar{y} = \dfrac{8}{4} = 0\!\!/\!\!/$

$S_{xy} = \dfrac{1}{4-1} \times (0+0+0+0) = 0\!\!/\!\!/$

$S_x = 2$ $S_y = 2$

$cor = \dfrac{0}{\sqrt{2} \times \sqrt{2}} = 0\!\!/\!\!/$

Euclidian

$d(x,y) = \sqrt{(0-1)^2 + (1-0)^2 + (0-1)^2 + (1-0)^2}$

$= 2\!\!/\!\!/$

Jaccard

$J = \dfrac{f_{11}}{f_{01} + f_{10} + f_{11}} = \dfrac{0}{2+2+0} = 0\!\!/\!\!/$

c) $x = (0, -1, 0, 1)$ $\qquad$ $y = (1, 0, -1, 0)$

Euclidean $\qquad$ cosine

dist.type $\qquad$ $x \cdot y = 0$ $\qquad$ $\|x\| = \sqrt{2}$ $\quad$ $\|y\| = \sqrt{2}$

$$\cos(x, y) = 0/2 = 0 \text{//}$$

correlation:

$$\bar{x} = 0/4 = 0\text{//} \qquad \bar{y} = 0/4 = 0\text{//}$$

$$Cov(x, y) = \frac{1}{4-1}(0) = 0\text{//}$$

$$corr = \frac{0}{0 \times 0} = 0\text{//}$$

Euclidean

$$d(x, y) = \sqrt{(0-1)^2 + (-1-0)^2 + (0+1)^2 + (1-0)^2}$$

$$= 2\text{//}$$

d) $x = (1, 1, 0, 1, 0, 1)$ $y = (1, 1, 1, 0, 0, 1)$

cosine

$x \cdot y = 1 \times 1 + 1 \times 1 + 0 \times 1 + 1 \times 0 + 0 \times 0 + 1 \times 1 = 3$

$\|x\| = \sqrt{4} = 2$ $\|y\| = \sqrt{4} = 2$

$\cos(x, y) = 3/4 = 0.75$

correlation

$\bar{x} = 4/6$ $\bar{y} = 4/6$

$cov(x \cdot y) = \frac{1}{(6-1)} \left[(1 - 4/6)(1 - 4/6) \cdots \right]$ (calc)

$= 1/15$

$SD(x) = 0.516$ $SD(y) = 0.516$

$cor(x, y) = \dfrac{1/15}{0.516 \times 0.516} = 0.25$

Jaccard

$f_{01} = 1$ $f_{10} = 1$ $f_{11} = 3$ $f_{00} = 1$

$J = \dfrac{3}{1 + 1 + 3} = 0.6$

e) $x = (2, -7, 0, 2, 0, -3)$ $y = (-1, 1, -1, 0, 0, -1)$

Cosine

$$x \cdot y = 0 \qquad \| x \| = \sqrt{18} \qquad \| y \| = 2$$

$$\cos(x, y) = \frac{0}{2 \times \sqrt{18}} = 0$$

Corelation

$$\text{cor}(x, y) = \frac{\text{cov}(x, y)}{SD(x) \times SD(y)}$$

x	y	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
2	-1	3	-2/3	-2	9	4/9
-1	1	-6	4/3	-20/3	36	16/9
0	-1	1	-2/3	-2/3	1	4/9
2	0	3	1/3	1	9	1/9
0	0	1	1/3	1/3	1	1/9
-3	-1	-2	-2/3	4/3	4	4/9

$$\bar{x} = -1 \qquad \bar{y} = -1/3 \qquad \Sigma = -20/3 \qquad \Sigma = 60 \qquad \Sigma = 1^{\text{c}}$$

$$\text{cor}(x, y) = \frac{-20/3}{60 \times 10/3} = \frac{-1}{30}$$

3) a) Sample Matching co-efficient

i) $SMC = \dfrac{f_{11} + f_{00}}{f_{01} + f_{10} + f_{11} + f_{00}}$

$f_{01} = 1$, $f_{10} = 2$, $f_{11} = 2$, $f_{00} = 5$

$SMC = \dfrac{2 + 5}{1 + 2 + 2 + 5} = \dfrac{7}{10} = 0.7$

ii) $f_{01} = 2$ $f_{10} = 1$ $f_{00} = 7$ $f_{11} = 0$

$SMC = \dfrac{0 + 7}{2 + 1 + 7 + 0} = 0.7$

b) Jaccord

i) $J = \dfrac{f_{11}}{f_{01} + f_{10} + f_{11}} = \dfrac{2}{1 + 2 + 2} = 0.4$

ii) $J = \dfrac{f_{11}}{f_{01} + f_{10} + f_{11}} = \dfrac{0}{2 + 1 + 0} = 0$

c) Hammering distance

i) $d(x,y) = |x_1 - y_1| + |x_2 - y_2| + \dots + |x_n - y_n|$

$= |0-0| + |1-1) + |0-0) + |1-0| + |0-0| + |7-7)$

$+ |0-1| + |6-0| + |0-0| + |1-0)$

$= \cancel{0+8} \; 3$

ii) $x = 1000000000$

$y = 000000100 1$

$= 1 + 0 + 0 + 0 + 0 + 0 + 1 + 0 + 0 + 1$

$= 3$

4) a) Cosine

$x \cdot y = 3 + 0 + 6 + 0 + 6 + 0 + 2 + 0 + 0$

$= 5$

$$\|x\| = \sqrt{42}$$

$$\|y\| = \sqrt{6}$$

$$\cos(x,y) = \frac{5}{\sqrt{42}\,\sqrt{6}}$$

$$= 0.\underline{\underline{314}}$$

b) correlation

~~cor (x,y)~~

x	y	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
3	1	1.8	0.6	1.08	3.82	0.36
2	0	0.8	-0.4	-0.32	0.64	0.16
0	0	-1.2	-0.4	0.48	1.44	0.16
5	0	3.8	-0.4	-1.52	14.44	0.16
0	0	-1.2	-0.4	0.48	1.44	0.16
6	0	-1.2	-0.4	0.48	1.44	0.16
0	0	-0.2	-0.4	0.12	0.04	0.36
2	1	0.2	0.6	0.48	1.44	0.16
0	0	-1.2	-0.4	-1.92	1.44	2.56
0	2	-1.2	1.6			
$\bar{x} = 1.2$	$\bar{y} = 0.4$			0.16	27.56	4.4

$$\text{cor}(x,y) = \frac{-0.16}{27.58 \times 4.4}$$

$$= -1.32 \times 10^{-3}$$