Para medir a dissemelhança entre 2 eventos de D vamos definir uma métrica uniforme. Por exemplo:

$$d_1(x,x') = \begin{cases} 1 & \text{se } x'_1 \neq x_1 \\ 0 & \text{se } x'_1 = x_1 \end{cases}, \quad d_2(x,x') = \begin{cases} 1 & \text{se } x'_2 \neq x_2 \\ 0 & \text{se } x'_2 = x_2 \end{cases}, \quad d_3(x,x') = \begin{cases} 1 & \text{se } x'_3 \neq x_3 \\ 0 & \text{se } x'_3 = x_3 \end{cases}.$$

e $\boxed{d_u(x,x') = \frac{1}{3}\left(d_1(x,x') + d_2(x,x') + d_3(x,x')\right)}$.

Neste caso, estamos a considerar a função $d_u(x,x')$ que dá igual peso à diferença entre cada um dos 3 atributos de dois eventos diferentes da base de dados D.

Começamos por provar que $d_u$ é uma função de dissemelhança. ($d(x,x') \geq 0 \wedge d(x,x) = 0$)

1º $d_u(x,y) \geq 0$

é fácil de ver que $d_u(x,y) = \frac{1}{3}d_1(x,x') + \frac{1}{3}d_2(x,x') + \frac{1}{3}d_3(x,x')$.

Como $d_1(x,x')$ ou é 0 ou é 1 e o mesmo acontece p/ $d_2(x,x')$ e $d_3(x,x')$, a sua soma será sempre entre 0 e 1.

2º $d_u(x,x) = 0$

$x = x'$ significa que $x_1 = x'_1 \wedge x_2 = x'_2 \wedge x_3 = x'_3$, logo $d_1(x,x') = 0$, $d_2(x,x') = 0$, $d_3(x,x') = 0$

pelo que $d_u(x,x') = \frac{1}{3}(0+0+0) = 0$

//

Também se verifica facilmente que goza a propriedade de simetria.

$\forall x, y \in \mathcal{A}, \; d_u(x,y) = d_u(y,x)$.

$$d_u(x,y) = \frac{1}{3}\left(d_1(x,y) + d_2(x,y) + d_3(x,y)\right)$$

onde $d_1(x,y) = \begin{cases} 1, & x_1 \neq y_1 \\ 0, & x_1 = y_1 \end{cases}$, $d_2(x,y) = \begin{cases} 1, & x_2 \neq y_2 \\ 0, & x_2 = y_2 \end{cases}$, $d_3(x,y) = \begin{cases} 1, & x_3 \neq y_3 \\ 0, & x_3 = y_3 \end{cases}$

por outro lado,

$$d_u(y,x) = \frac{1}{3}\left(d_1(y,x) + d_2(y,x) + d_3(y,x)\right)$$

onde $d_1(y,x) = \begin{cases} 1, & y_1 \neq x_1 \\ 0, & y_1 = x_1 \end{cases}$, $d_2(y,x) = \begin{cases} 1, & y_2 \neq x_2 \\ 0, & y_2 = x_2 \end{cases}$, $d_3(y,x) = \begin{cases} 1, & y_3 \neq x_3 \\ 0, & y_3 = x_3 \end{cases}$.

As funções $d_1(x,y) = d_1(y,x)$ porque quando $x_1 \neq y_1$ também $y_1 \neq x_1$ e quando $x_1 = y_1$ também $y_1 = x_1$. O mesmo se verifica para $d_2$ e $d_3$.

$\therefore$ $d_u$ é uma função de dissemelhança simétrica.

- Também verifica definiteness.

$\forall x, y \in \mathcal{A}, \; d_u(x,y) = 0 \Leftrightarrow x = y$.

Notar que $d_u(x,y) = 0 \iff \frac{1}{3}\left(d_1(x,y) + d_2(x,y) + d_3(x,y)\right) = 0$.

$d_u(x,y)=0 \rightarrow x=y$

como $d_1, d_2$ e $d_3 \geqslant 0$, $d_u$ só pode ser zero se $d_1(x,y)=d_2(x,y)=d_3(x,y)=0$.

$$\left.\begin{array}{ll} d_1(x,y)=0 & \text{então } x_1=y_1 \\ d_2(x,y)=0 & \text{então } x_2=y_2 \\ d_3(x,y)=0 & \text{então } x_3=y_3 \end{array}\right\} \quad x=y.$$

2) $x=y \rightarrow d_u(x,y)=0$

$x=y \Leftrightarrow x_1=y_1 \wedge x_2=y_2 \wedge x_3=y_3$.

$$d_u(x,y)=\frac{1}{3}\Big(\underbrace{d_1(x,y)}_{0}+\underbrace{d_2(x,y)}_{0}+\underbrace{d_3(x,y)}_{0}\Big)$$

∴ $d_u$ é uma função dissemelhança simétrica e que goza da propriedade de definiteness.

- Finalmente, vamos verificar se a nossa proposta (dissemelhança) verifica a desigualdade triangular, e como tal é uma dissemelhança métrica (distância)
- Recordar que para que tal aconteça, $\boxed{\forall x,y,z \in \mathcal{X},\ d(x,z) \leq d(x,y)+d(y,z)}$

$d_u(x,z) = \frac{1}{3}\left(d_1(x,z) + d_2(x,z) + d_3(x,z)\right)$

$d_u(x,y) = \frac{1}{3}\left(d_1(x,y) + d_2(x,y) + d_3(x,y)\right)$

$d_u(y,z) = \frac{1}{3}\left(d_1(y,z) + d_2(y,z) + d_3(y,z)\right)$

então, usando a def de desigualdade triangular, vem,

$$\frac{1}{3}\left(d_1(x,z)+d_2(x,z)+d_3(x,z)\right) \leq \frac{1}{3}\left(d_1(x,y)+d_2(x,y)+d_3(x,y)\right) + \frac{1}{3}\left(d_1(y,z)+d_2(y,z)+d_3(y,z)\right) \Leftrightarrow$$

$$\Leftrightarrow d_1(x,z) + d_2(x,z) + d_3(x,z) \leq d_1(x,y) + d_1(y,z) + d_2(x,y) + d_2(y,z) + d_3(x,y) + d_3(y,z)$$

Já vimos antes que $d_i(x,y) \geq 0$. Por outro lado,

$d_1(x,z) = \begin{cases} 1, & x_1 \neq z_1 \\ 0, & x_1 = z_1 \end{cases}$, $d_2(x,z) = \begin{cases} 1, & x_2 \neq z_2 \\ 0, & x_2 = z_2 \end{cases}$, $d_3(x,z) = \begin{cases} 1, & x_3 \neq z_3 \\ 0, & x_3 = z_3 \end{cases}$

- Se $d_1(x,z) = 0$, $0 \leq \underbrace{d_1(x,y)}_{[0,1]} + \underbrace{d_1(y,z)}_{[0,1]}$

- se $d_1(x,z) \neq 0$ ou seja $d_1(x,z) = 1$ significa que $x_1 \neq z_1$. A desigualdade só não se verifica se $d_1(x,y) = 0 \wedge d_1(y,z) = 0$

$\begin{cases} d_1(x,y) = 0 & \text{então } x_1 = y_1 \\ d_1(y,z) = 0 & \text{então } y_1 = z_1 \end{cases} \rightarrow x_1 = z_1$ o que é uma contradição, logo não é possível $d_1(x,z) = 1 \wedge d_1(y,z) + d_1(x,y) = 0$ (redução ao absurdo)

- Da mesma forma se prova que $d_2(x,z) \leq d_2(x,y) + d_2(y,z)$ e $d_3(x,z) \leq d_3(x,y) + d_3(y,z)$. $d_u(x,y)$ é uma dessemelhança

## DADOS BINÁRIOS

10) $X \in \mathcal{X} = A_1 \times A_2 \times \dots \times A_I = \{0,1\}^I$

$A_i = \{0,1\}$

Vamos considerar o exemplo que consiste em quantificar a semelhança entre duas imagens de 4 pixels.

| | |
|---|---|
| 1 | 2 |
| 3 | 4 |

$X = (x_1, x_2, x_3, x_4)$

$X \in \mathcal{X} = \{0,1\}^4$

Para tal, vamos definir uma função de semelhança pixel a pixel,

$$s(x,x') = \frac{1}{4}\left(s_1(x,x') + s_2(x,x') + s_3(x,x') + s_4(x,x')\right)$$

onde,

$$s_1(x,x') = \begin{cases} 1, & x_1 = x'_1 \\ 0, & x_1 \neq x'_1 \end{cases}, \quad s_2(x,x') = \begin{cases} 1, & x_2 = x'_2 \\ 0, & x_2 \neq x'_2 \end{cases}, \quad s_3(x,x') = \begin{cases} 1, & x_3 = x'_3 \\ 0, & x_3 \neq x'_3 \end{cases}, \quad s_4(x,x') = \begin{cases} 1, & x_4 = x'_4 \\ 0, & x_4 \neq x'_4 \end{cases}$$

Provar que $s(x,x')$ é uma função de semelhança.

- $s(x,y) \in \mathbb{R}$
- $s(x,x) \geq s(x,y)$
- $s(x,y) = s(y,x)$

- $s(x,x')$ é a soma de 4 funções cujo output é 0 ou $1^{(\in \mathbb{R})}$, a soma de 4 n.ºs reais é ainda um n.º real.

- $$s(x,x) = \frac{1}{4}\left(s_1(x,x) + s_2(x,x) + s_3(x,x) + s_4(x,x)\right) = 1$$

$$s(x,y) = \frac{1}{4}\left(\underset{0/1}{s_1(x,y)} + \underset{0/1}{s_2(x,y)} + \underset{0/1}{s_3(x,y)} + \underset{0/1}{s_4(x,y)}\right) \in [0,1]$$

$$s(y,x) = \frac{1}{4}\left(\underset{0/1}{s_1(y,x)} + \underset{0/1}{s_2(y,x)} + \underset{0/1}{s_3(y,x)} + \underset{0/1}{s_4(y,x)}\right) \in [0,1]$$

$\therefore$ $s(x,y)$ é uma função de semelhança.

Ex11)

- $s(x,y)$ é simétrica,

$$s(x,y) = \frac{1}{4}\left(s_1(x,y) + s_2(x,y) + s_3(x,y) + s_4(x,y)\right)$$

$$s(y,x) = \frac{1}{4}\left(s_1(y,x) + s_2(y,x) + s_3(y,x) + s_4(y,x)\right)$$

$$s_1(x,y) = \begin{cases} 1, & x_1 = y_1 \\ 0, & x_1 \neq y_1 \end{cases} \qquad s_1(y,x) = \begin{cases} 1, & y_1 = x_1 \\ 0, & y_1 \neq x_1 \end{cases}$$

Quando $x_1 = y_1$, também $y_1 = x_1$, logo $s_1(x,y) = s_1(y,x)$.
Tal tb acontece para $s_2(x,y) = s_2(y,x)$, $s_3(x,y) = s_3(y,x)$ e $s_4(x,y) = s_4(y,x)$.

$\therefore$ $s(x,y) = s(y,x)$ pelo que $s$ é uma função de semelhança simétrica.

- $s(x,y)$ é normalizada pq $s_1(x,y) \in [0,1]$, $s_2(x,y) \in [0,1]$, $s_3(x,y) \in [0,1]$, $s_4(x,y) \in [0,1]$

$\therefore \frac{1}{4}\left(s_1(x,y)+s_2(x,y)+s_3(x,y)+s_4(x,y)\right) \in [0,1]$.

Ex12

- $s(x,y)$ verifica definiteness.

$\forall x,y \in \mathcal{A}, \; s(x,y)=1 \Leftrightarrow x=y$.

- $s(x,y)=1 \rightarrow x=y$

$s(x,y)=1 \Leftrightarrow \frac{1}{4}\left(s_1(x,y)+s_2(x,y)+s_3(x,y)+s_4(x,y)\right) = 1 \Leftrightarrow$

$\Leftrightarrow s_1(x,y)+s_2(x,y)+s_3(x,y)+s_4(x,y) = 4$.

Como $s_i(x,y) \in [0,1]$, a única possibilidade é $s_1(x,y)=1 \wedge s_2(x,y)=1 \wedge s_3(x,y)=1 \wedge s_4(x,y)=1$.

Para $s_1(x,y)=1$ temos que $x_1=y_1$
$s_2(x,y)=1$ " " $x_2=y_2$
$s_3(x,y)=1$ " " $x_3=y_3$
$s_4(x,y)=1$ " " $x_4=y_4$

} se tudo isto acontecer em simultâneo é pq $x=y$.

- $x=y \rightarrow s(x,y)=1$

$x=y$ significa que $x_1=y_1 \wedge x_2=y_2 \wedge x_3=y_3 \wedge y_4=x_4$.

Daqui vem que $s_1(x,y)=1$ ; $s_2(x,y)=1$ ; $s_3(x,y)=1$ e $s_4(x,y)=1$.

$\therefore s(x,y) = \frac{1}{4}(1+1+1+1) = 1$

//

**Monoide** (grupóide associativo c/ elemento neutro)

- Seja o espaço dos atributos $\mathcal{A} = A \times A \times A \times \cdots \times A = A^I$
  onde $A = \{a, b, c, \ldots, z, \varepsilon\}$.
  elemento neutro

Ex16
- Façamos $I = 5$ (por exemplo).
  Estamos a considerar palavras até 5 letras.

- Uma função dissemelhança poderia ser:

  $\boxed{d(x, x') = \text{número de letras de } x \text{ que são diferentes de } x'.}$

  1º $d(x, x') \geq 0$ – verdade pq se são todas iguais dá 0, se há letras de x diferentes, ou output de d é zero!

  2º $d(x, x) = 0$

  $\therefore$ d é dissemelhança

- d não é simétrica – contra exemplo

$d(\text{'aula'}, \text{'aulas'}) = 0$
$d(\text{'aulas'}, \text{'aula'}) = 1$ $\longrightarrow d(x, x') \neq d(x', x)$

- d não está normalizada

$d(\text{'sol'}, \text{'rua'}) = 3 \notin [0, 1]$

- d não verifica definiteness

$d(x, x') = 0 \Leftrightarrow x = x'$

$d(\text{'aula'}, \text{'lua'}) = 0 \wedge \text{'aula'} \neq \text{'lua'}$

Outras funções $d(x, x')$ podem ser construídas.
Vamos agora considerar exemplo e definir uma função semelhança.

$\mathcal{A} = A \times A \times A \times \dots \times A = A^5$
$A = \{a, b, c, \dots, z, \varepsilon\}$

$$\boxed{\begin{array}{l} s(x, x') = \frac{1}{5}\left(s_1(x, x') + s_2(x, x') + s_3(x, x') + s_4(x, x') + s_5(x, x')\right) \\ \text{onde } s_1(x, x') = \begin{cases} 1, & x_1 = x'_1 \\ 0, & x_1 \neq x'_1 \end{cases} ; \; s_2(x, x') = \begin{cases} 1, & x_2 = x'_2 \\ 0, & x_2 \neq x'_2 \end{cases} \dots \; s_i(x, x') = \begin{cases} 1, & x_i = x'_i \\ 0, & x_i \neq x'_i \end{cases} \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = 1, \dots 5 \end{array}}$$