Métricas em Machine Learning

2. Métricas ponto a ponto

- Em ML os algoritmos são baseados na quantificação de semelhança ou dissemilhança de dados;
- De acordo com os tipos de dados e os problemas, há métricas mais adequadas que outras.
- Em problemas de clustering (essencialmente os que vamos tratar) precisamos de quantificar a proximidade entre 2 eventos numa BD.

Assim, considere-se que o espaço de atributos $\mathcal{A} = A_1 \times A_2 \times ... \times A_I$.
Sejam $x = (x_1, x_2, ..., x_I) \in \mathcal{A}$ e $x' = (x'_1, x'_2, ..., x'_I) \in \mathcal{A}$

Definição: d é uma função de dissemelhança se:

i) $d(x, x') \geq 0$ (não negatividade).
ii) $d(x, x) = 0$ (reflexividade).

Esta função pretende dizer qual o grau de afastamento entre x e x'.

Exemplo 7: Considere a função,

$$d(x,y) = \begin{cases} 0 & se\ x_1 = y_1 \\ |x_2 - y_2| & se\ x_1 \neq y_1 \end{cases}$$

Verifique que é uma dissemelhança. Calcule $d(x,y)$ para $x = (sim, 0.7)$ e $y = (Nao, 0.25)$ com $x, y \in \mathcal{A}$ com $\mathcal{A} = \{sim, nao\} \times [0,1]$.

Solução:

Dissemelhança pq $d(x,y)$ só tem 2 outputs possíveis 0 ou o módulo dum número, logo é ≥ 0. $d(x,x) = 0$ pq se $y = x$, então $x_1 = y_1 \rightarrow d(x,x) = 0$.

$x = (sim, 0.7)$ e $y = (Nao, 0.25) \rightarrow x_1 \neq y_1 \rightarrow d(x,y) = |0.7 - 0.25| = 0.45$.

Métricas em Machine Learning

Definição: Uma função de dissemelhança d diz-se simétrica se,

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

Definição: Uma função de dissemelhança d verifica *definitness* (identidade de indiscerníveis) se,

$\forall x, y \in \mathcal{A},\ d(x, y) = 0 \equiv x = y$.

Definição: Uma função de dissemelhança d que verifique,

$\forall x, y, z \in \mathcal{A},\ d(x, z) \leq d(x, y) + d(y, z)$,
diz-se que verifica a desigualdade triangular.

Definição: Uma função d que verifique,

a) $d(x, y) \geq 0$ e $d(x, x) = 0$
b) $d(x, y) = d(y, x)$.
c) $\forall x, y \in \mathcal{A},\ d(x, y) = 0 \equiv x = y$. .
d) $\forall x, y, z \in \mathcal{A},\ d(x, z) \leq d(x, y) + d(y, z)$,
diz-se uma dissemelhança métrica (distância).

Métricas em Machine Learning

Até agora considerámos $D = \{e^n, n = 1, ...M\}$ e o evento $e^n = (x^n, y^n)$ com $x^n \in \mathcal{A}$ e $y^n \in \mathcal{C}$. Como vamos essencialmente fazer *clustering* (UML), vamos a partir de agora dizer $e^n = x^n$.

Dados nominativos

Considere-se que temos uma base de dados D que para caracterizar os frutos banana, laranja, limão, maça e amêndoa,

$A_1 = cor = \{amarelo, laranja, verde, vermelho, castanho\}$

$A_2 = sabor = \{doce, amargo, acido\}$

$A_3 = forma = \{alongado, redondo\}$

O espaço dos atributos $\mathcal{A} = A_1 \times A_2 \times A_3$.

Um evento da base de dados D: $x^1 = (amarelo, doce, alongado)$.
Notação para descrever o evento $e^n = (x_1^n, x_2^n, x_3^n)$ com $x_1 \in A_1$, $x_2 \in A_2$ e $x_3 \in A_3$.

Métricas em Machine Learning

Para medir uma dissemelhança entre 2 eventos de D podemos definir a **métrica** uniforme d_u tal que,

$$d1(x,x') = \begin{cases} 1 & se\ x_1 \neq x'_1 \\ 0 & se\ x_1 = x'_1 \end{cases}$$

$$d2(x,x') = \begin{cases} 1 & se\ x_2 \neq x'_2 \\ 0 & se\ x_2 = x'_2 \end{cases}$$

$$d3(x,x') = \begin{cases} 1 & se\ x_3 \neq x'_3 \\ 0 & se\ x_3 = x'_3 \end{cases}$$

e,

$$d_u(x,x') = \tfrac{1}{3}(d1(x,x') + d2(x,x') + d3(x,x')).$$

Exemplo 8: Mostrar que $d_u(x,x')$ é uma dissemelhança métrica.

Solução:

1. Mostrar de $d_u(x, y)$ é função de dissemelhança.

a) $d_u(x, y) \geq 0$ ou seja $d_u(x, x') = \frac{1}{3}(d1(x, x') + d2(x, x') + d3(x, x')) \geq 0$. Uma vez que $d_1(x, y)$ ou vale 0 ou vale 1 e o mesmo acontece com $d_2(x, y)$ e $d_3(x, y)$, então, efetivamente $d_u(x, y) \geq 0$.

b) $d(x, x) = 0$. Vemos que $y = x$, ou seja $x_1 = y_1$, $x_2 = y_2$, e $x_3 = y_3$. Ou seja $d1(x, y) = 0$, $d2(x, y) = 0$ e $d3(x, y) = 0$, logo, $d(x, y) = 0$ quando $x = y$.

2. Mostrar de $d_u(x, y)$ goza da simetria $(d_u(x, y) = d_u(y, x))$

Vimos que $d_u(x, y) = \frac{1}{3}(d1(x, y) + d2(x, y) + d3(x, y))$. Por outro lado, $d_u(y, x) = \frac{1}{3}(d1(y, x) + d2(y, x) + d3(y, x))$.

Vemos que $d1(x, y) = d1(y, x)$ uma vez que quando $x_1 \neq y1$ também $y_1 \neq x1$. O mesmo se verifica para $d_2(x, y)$ e $d3(x, y)$.

3. Mostrar *definitness*

$\forall x, y \in \mathcal{A} : d_u(x, y) = 0 \equiv x = y$. Temos que provar os 2 lados da implicação. A saber,

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

Como $d1, d2, d3 \geq 0$, d_u sópode ser 0 se $d1(x, y) = d2(x, y) = d3(x, y) = 0$.

$d1(x, y) = 0 \rightarrow x_1 = y1$

$d2(x, y) = 0 \rightarrow x_2 = y2$

$d3(x, y) = 0 \rightarrow x_3 = y3$

Então, $x = y$.

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

$x = y$ é o mesmo que $x_1 = y_1 \wedge x_2 = y_2 \wedge x_3 = y_3$. Então
$d1(x, y) = d2(x, y) = d3(x, y) = 0$.

4. Mostrar de $d_u(x, y)$ goza da desigualdade trangular
$(\forall x, y, z \in \mathcal{A},\ d_u(x, z) \leq d_u(x, y) + d_u(y, z)\ )$

[TPC]

Notar, que podemos definir outras funções de dissemelhança. Por exemplo poderímos atribuir pesos diferentes às várias parcelas.

$d_u(x, y) = w1.d1(x, y) + w2.d2(x, y) + w3.d3(x, y)$ onde $w1 + w2 + w3 = 1$ e $w1 \geq 0, w2 \geq 0, w3 \geq 0$.

Métricas em Machine Learning

Definição : s é uma função de semelhança se:

a) $s(x, y) \in \mathbb{R}$;
b) $s(x, x) \geq s(x, y)$ e $s(x, x) \geq s(y, x)$.

s é uma função que mede a semelhança entre dois dados x e $y \in \mathcal{A}$ (espaço dos atributos).

Definição : s é uma semelhança simétrica se:

a) $s(x, y) \in \mathbb{R}$;
b) $s(x, x) \geq s(x, y)$.
c) $s(x, y) = s(y, x)$.

Definição : s é uma semelhança simétrica e normalizada se:

a) $s(x, y) \in [0, 1]$;
b) $s(x, x) \geq s(x, y)$.
c) $s(x, y) = s(y, x)$.
Quando os atributos têm ordens de grandeza muito diferentes, é costume normalizar.

Métricas em Machine Learning

Definição: Uma função s é uma função de semelhança, semétrica, normalizada e que verifica *definitness* se,

a) $s(x, y) \in [0, 1]$;
b) $s(x, x) \geq s(x, y)$.
c) $s(x, y) = s(y, x)$.
d) $s(x, y) = 1 \equiv x = y$.

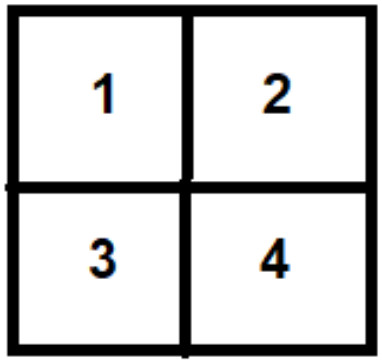
Dados binários

$x \in \mathcal{A} = A_1 \times A_2 \times ... \times A_I = \{0,1\}^I$

Uma das métricas mais usadas com dados binários é a distância de Hamming.

Para comprrender métricas que se possam definir, associadas a dados binários, vamos ter por base uma exemplo.

Exemplo 9: Quantificar a semelhança entre 2 imagens de 4 pixels, que podem ser brancos ou pretos - $x \in \mathcal{A} = A_1 \times A_2 \times A_3 \times A_4 = \{0,1\}^4$.



Métricas em Machine Learning

Vamos definir uma função semelhança pixel a pixel da seguinte forma,

$s(x, x') = \frac{1}{4}(s1(x, x') + s2(x, x') + s3(x, x') + s4(x, x'))$, onde,

$$s1(x, x') = \begin{cases} 0 & se\ x_1 \neq x'_1 \\ 1 & se\ x_1 = x'_1 \end{cases}$$

$$s2(x, x') = \begin{cases} 0 & se\ x_2 \neq x'_2 \\ 1 & se\ x_2 = x'_2 \end{cases}$$

$$s3(x, x') = \begin{cases} 0 & se\ x_3 \neq x'_3 \\ 1 & se\ x_3 = x'_3 \end{cases}$$

$$s4(x, x') = \begin{cases} 0 & se\ x_4 \neq x'_4 \\ 1 & se\ x_4 = x'_4 \end{cases}$$

Exemplo 10 : Mostrar que s é uma função de semelhança.

Solução: Temos que mostrar 3 coisas,
1) $s(x, y) \in \mathbb{R}$.
2) $s(x, x) \geq s(x, y)$.
3) $s(x, x) \geq s(y, x)$.

- $s(x, x')$ é a soma de 4 funções cujo valor é 0 ou 1. A soma de 4 números reais é um número real.

- $s(x, x) = \frac{1}{4}(s1(x, x) + s2(x, x) + s3(x, x) + s4(x, x)) = 1$;

$$s(x, x') = \frac{1}{4}(s1(x, x') + s2(x, x') + s3(x, x') + s4(x, x')) = \\ = \frac{1}{4}.((0 \ ou \ 1) + (0 \ ou \ 1) + (0 \ ou \ 1) + (0 \ ou \ 1))$$

então, $s(x, x) \geq s(x, x')$.

Exemplo 11: Mostrar que s é simétrica.

Pretende-se mostrar que $s(x, y) = s(y, x)$

$s(x, y) = \frac{1}{4}(s1(x, y) + s2(x, y) + s3(x, y) + s4(x, y)$ e
$s(y, x) = \frac{1}{4}(s1(y, x) + s2(y, x) + s3(y, x) + s4(y, x)$.

Olhemos para uma das parcelas das expressões acima.

$$s1(x, y) = \begin{cases} 0 & se \ x_1 \neq y_1 \\ 1 & se \ x_1 = y_1 \end{cases} \quad s1(y, x) = \begin{cases} 0 & se \ y_1 \neq x_1 \\ 1 & se \ y_1 = x_1 \end{cases}$$

Quando $x_1 = y_1$, também $y_1 = x_1$, logo $s1(x, y) = s1(y, x)$ pelo que $s(x, y) = s(y, x)$.

Exemplo 12: Mostrar que $s(x, y)$ é normalizada e verifica a propriedade de *definitness*.

Matriz de confusão para dois eventos binários - $x, y \in \{0, 1\}^I$

Na aula passada já vimos o que é a matriz de confusão. Recordemos,

D \ D'	0	1
0	M00	M01
1	M10	M11

$M00(x, y) = \sum_{i=1}^{I} \neg x_i \neg y_i$ - número de vezes que x e y têm 0 em comum.

$M01(x, y) = \sum_{i=1}^{I} \neg x_i y_i$ - número de vezes que x é 0 e e y é 1.

$M10(x, y) = \sum_{i=1}^{I} x_i \neg y_i$ - número de vezes que x é 1 e e y é 0.

Métricas em Machine Learning

$M11(x, y) = \sum_{i=1}^{I} x_i y_i$ - número de vezes que x e y têm 1 em comum.

A partir de $M00$, $M01$, $M10$ e $M11$ podemos definir funções de dissemelhança entre dois eventos de dados x e y. Por outro lado, $M00 + M01 + M10 + M11 = I$ que é o número de atributos dum evento x.

Podemos definir,

$$d(x, y) = \frac{M01 + M10}{I}.$$

Exemplo 13 : Mostar que $d(x, y)$ definido antes é uma dissemelhança.

Temos que provar $d(x, x) = 0$ e que $0 \leq d(x, y) \leq 1$.

Exemplo 14 : Mostar que $d(x, y)$ definido antes é uma dissemelhança simétrica.

Pista: Pode-se usar o facto que $M10(x, y) = M01(y, x)$ e ainda que $M10(y, x) = M01(x, y)$.

Métricas em Machine Learning

A seguir apresenta-se um conjunto de métricas de semelhança e dissemelhança usadas no contexto de eventos usando dados binários.

Semelhança de Jaccard : $s_J = \frac{M11}{M01 + M10 + M11}$.

Dissemelhança de Jaccard : $d_J = \frac{M01 + M10}{M01 + M10 + M11}$.

Semelhança de Dice : $s_D = \frac{2M11}{M01 + M10 + 2M11}$.

Dissemelhança de Dice : $d_D = \frac{M01 + M10}{M01 + M10 + 2M11}$.

Semelhança Overlap : $s_O = \frac{M11 + M00}{M01 + M10 + M11 + M00}$.

Disemelhança Overlap : $d_O = \frac{M01 + M10}{M01 + M10 + M11 + M00}$.

A métrica a escolher depende da natureza do problema.

Exemplo 15: Considere $x = (1, 0, 1, 0, 0, 0)$ e $y = (1, 1, 1, 0, 0, 0)$

a) Avaliar a semelhança e dissemelhança entre x e y usando Jaccard.
b) Avaliar a semelhança e dissemelhança entre x e y usando Overlap.