Exercises – PART I

KNN

1. 3-NN (1.0, 0.5) → (P)

±0,25

NN: (0,75 ; 0,25) → $d = \sqrt{0,25^2 + 0,25^2} = 0,3536$ (N)

(0,75 ; 0,75) → $d = \sqrt{0,25^2 + 0,25^2} = 0,3536$ (N)

(1,25 ; 0,25) → $d = \sqrt{0,25^2 + 0,25^2} = 0,3536$ (N)

(1,25 ; 0,75) → $d = \sqrt{0,25^2 + 0,25^2} = 0,3536$ (P)

(C) is not able to classify

2. 3-NN (1,75 ; 2)

→ (2,2) → $d = \sqrt{0,25^2} = 0,25$ [P]

→ (1,5 ; 2) → $d = \sqrt{0,25^2} = 0,25$ [N]

(B)

- (1,2) $d = \sqrt{0,75^2 + 0} = 0,75$ [N]
- (1,75 ; 0) $d = \sqrt{0 + 2^2} = 2$
- (1,25 ; 3) $d = \sqrt{0,5^2 + 1^2}$

3. (C)

4.

x1	x2	x3	clam
0	0	1	0
0	1	1	[1]
1	0	1	[1]
1	1	1	[0] ←
0	0	0	0
0	1	0	0
1	0	0	0
1	1	0	[0]

a) (1,1,1) 3NN (B)

$d = \sqrt{1^2 + 0 + 0} = 1$

$d = \sqrt{0 + 1^2 + 0} = 1$

$d = \sqrt{0 + 0 + 1^2} = 1$

b) Leave-one-out

x1	x2	x3	clam	
0	0	0	0	TN
0	0	1	0	TN
0	1	0	0	TN
0	1	1	0	FN
1	0	0	0	TN
1	0	1	0	FN
1	1	0	1	FP
1	1	1	1	FP

$\text{Accuracy} = \frac{TP + TN}{All} = \frac{4}{8} = \frac{1}{2}$

Naive Bayes

1.

a) instance (A,B)

$P(C=P \mid (A,B)) = P(C=P) \times P(A \mid P) \times P(B \mid P) =$

$= \frac{3}{8} \times 1 \times \frac{1}{3} = \frac{1}{8}$ (B) Negative

$P(C=N \mid (A,B)) = P(C=N) \times P(A \mid N) \times P(B \mid N) =$

$= \frac{5}{8} \times \frac{2}{5} \times \frac{4}{5} = \frac{1}{5} >$

2. instance (T,T,F) $P(P) = 2/8 = 1/4$ $P(N) = 3/4$

$P(C=P \mid (T,T,F)) = P(C=P) \times P(A=T \mid P) \times P(B=T \mid P) \times P(C=F \mid P) =$

$= \frac{1}{4} \times \frac{2}{5} \times 0 = 0$ (B) Negative

$P(C=N \mid (T,T,F)) > 0$

4. X1 numerical e X2 ordinal instance (0.75; 1)

$P(C=P \mid (0.75;1)) = P(C=P) \times P(X1=0,75 \mid P) \times P(X2=1 \mid P)$

$\downarrow \frac{7}{12}$ $\downarrow$ $\downarrow \frac{2}{7}$

$f(x_1 = 0,75)$

$\downarrow$

$\frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$

Decision Trees

1. A, B, C numerical

a) Entropy do atributo class

$E(P,N) = -P(P) \cdot \log_2 P(P) - P(N) \cdot \log_2 P(N) =$

$= -\frac{3}{5} \log_2 \left(\frac{3}{5}\right) - \frac{2}{5} \log_2 \left(\frac{2}{5}\right) \simeq 0,97$

b) Attribute tested in the root -> C4.5 (Information gain criteria)

$G(A) = E(P,N) - \%.0 \times E(0+,0-) - \%.1 \times E(1+,1-)$

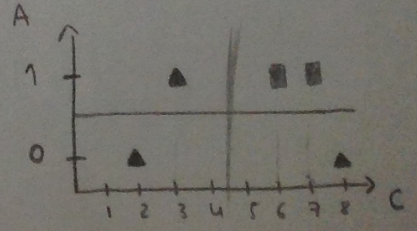
$= 0,97 - \frac{2}{5} \times E(2,0) - \frac{3}{5} \times E(1,2) = 0,42$

$E(2,0) = -\frac{2}{2} \log_2 \frac{2}{2} - 0 = -1 \times 0 = 0$

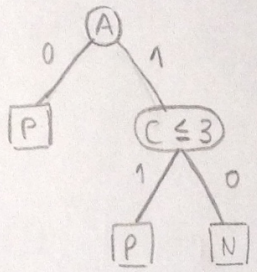
$E(1,2) = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} = 0,92$

↓ fazer as contas p/ os outros

$G(A) = G(C)$



c)



remar linhas c/ A = Ø

E(clone) = E(1,2) ≈ 0,92

G(B) = ... } o que dar > vai para o ramo

G(C) = ...

Ex. 8

a) $X2 \leq 1,25$ $X1 \leq 2,75$

$gini_{split}(X1 \leq 2,75) = P(X1 \leq 2,75) \times gini(X1 \leq 2,75) + P(X1 > 2,75) \times gini(X1 > 2,75) =$

$= \frac{8}{20} \times 0 + \frac{12}{20} \times 0,49 = 0,29$

$Gini(X1 \leq 2,75) = 1 - \left(\frac{8^2}{8^2}\right) = 0$

$Gini(X1 > 2,75) = 1 - \left(\frac{5^2 + 7^2}{12^2}\right) = 0,49$

(5² + 7²: quantos não P e não > 2,75; quantos não N e não > 2,75; 12²: total dos > 2,75)

$gini_{split}(X2 \leq 1,25) = P(X2 \leq 1,25) \times gini(X2 \leq 1,25) + P(X2 > 1,25) \times gini(X2 > 1,25) =$

$= \frac{10}{20} \times 0 + \frac{10}{20} \times 0,5 = \boxed{0,25}$

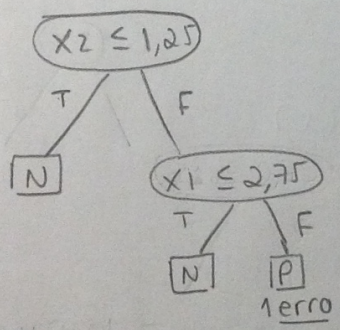
$Gini(X2 \leq 1,25) = 1 - \left(\frac{10^2}{10^2}\right) = 0$

$Gini(X2 > 1,25) = 1 - \left(\frac{5^2 + 5^2}{10^2}\right) = 0,5$

$\Rightarrow$ escolher gini < : 0,25 $\rightarrow$ X2

gini split < X2 ≤ 1,75 / X1 ≤ 2,75

b)



c) $\text{Accuracy} = \frac{TP+TN}{All} = \frac{19}{20}$

$\text{Sensibility} = \frac{TP}{TP+FN} = \frac{5}{5+0} = 1$

$\text{Specificity} = \frac{TN}{TN+FP} = \frac{14}{15}$

Ex.12 A and B Boolean , C numerical —— 0,25 ; 0,5 , 0,75 (values)

a) (A)

b) (A)

c) ENTROPY

$$E(P,N) = -\frac{8}{12}\log_2\frac{8}{12} - \frac{4}{12}\log_2\frac{4}{12}$$

$$= 0{,}918 \quad (C)$$

A	B	C	class
0	0	0,25	N ←
0	1	0,25	P
1	0	0,25	P
1	1	0,25	N ←
0	0	0,5	N ←
0	1	0,5	P
1	0	0,5	P
1	1	0,5	N ←
0	0	0,75	N ←
0	1	0,75	P
1	0	0,75	P
1	1	0,75	N ←

d) ACCURACY $= \frac{TP+TN}{All} = \frac{6+4}{12} = \frac{10}{12} = 0{,}83$

83% (B)

e) SENSIBILITY $= \frac{TP}{TP+FN} = \frac{6}{8} = 0{,}75$

SPECIFICITY $= \frac{TN}{TN+FP} = \frac{4}{4} = 1$

(B)

Ex.13. A, B, C numeric

a) (D)

b) (B)

c) ENTROPY $= -\frac{9}{27}\log_2\frac{9}{27} - \frac{18}{27}\log_2\frac{18}{27} = 0{,}764$ (B)

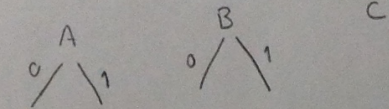
d) ACCURACY $= \frac{TP+TN}{All} = \frac{9+15}{27} = 0{,}89$ (B)

e) SENSIBILITY $= \frac{TP}{TP+FN} = \frac{6}{9} = 0{,}67$ (C)

SPECIFICITY $= \frac{TN}{TN+FP} = \frac{18}{18} = 1$

Ensembles

Ex.1. weak classifiers : all decision tree with only one binary node

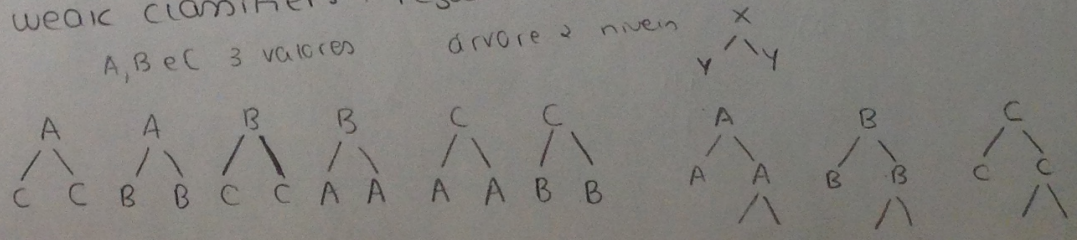


a) B

b) C

Ex.2. weak classifiers : regular trees with just two levels of nodes

A, B e C 3 valores árvore 2 níveis



ASSOCIATION RULES

Ex.2. ABCDF, ABEG, AD, BCDE, CD min support = 40%.

a)

C1:

A	3/5
B	3/5
C	3/5
D	4/5
E	2/5
~~F~~	~~1/5~~

L1:

A
B
C
D
E

C2:

AB	2/5
~~AC~~	~~1/5~~
AD	2/5
~~AE~~	~~1/5~~
BC	2/5
BD	2/5
BE	2/5
CD	3/5
~~CE~~	~~1/5~~
~~DE~~	~~1/5~~

L2:

AB
AD
BC
BD
BE
CD

C3:

ABD	
BCD	(A)
BCE	
BDE	

b)

CONFIDENCE $(A \to B) = P(B|A)$ SUPPORT $(A \to B) = P(A \cap B)$

LIFT $(A \to B) = \frac{P(B|A)}{P(B)}$

$$CONF(A \to D) = P(D|A) = \frac{P(D \cap A)}{P(A)} = \frac{\frac{2}{5}}{\frac{3}{5}} = \frac{2}{3}$$

$$CONF(D \to A) = P(A|D) = \frac{P(A \cap D)}{P(D)} = \frac{\frac{2}{5}}{\frac{4}{5}} = \frac{2}{4} = \frac{1}{2}$$

$$LIFT(A \to D) = \frac{P(D|A)}{P(D)} = \frac{P(D \cap A)}{P(D) \times P(A)} = \frac{\frac{2}{5}}{\frac{4}{5} \times \frac{3}{5}} = \frac{\frac{2}{5}}{\frac{12}{25}} = \frac{50}{60} = \frac{5}{6}$$

$$LIFT(D \to A) = \frac{CONF(D \to A)}{P(A)} = \frac{\frac{1}{2}}{\frac{2}{5}} = \frac{5}{6}$$ (A)

Ex.1.

a) min support = 40%.

C1:

A	2/4
B	3/4
C	3/4
~~D~~	~~1/4~~
E	3/4

L1:

A
B
C
E

C2:

~~AB~~	~~1/4~~
AC	2/4
~~AE~~	~~1/4~~
BC	2/4
BE	3/4
CE	2/4

L2:

AC
BC
BE
CE

C3: BCE 2/4

L3: BCE

C4: ∅

C: candidates

L: frequent item sets

$$CONF(X \to Y) = P(Y|X) = \frac{P(Y \cap X)}{P(X)}$$

b) ASSOC. RULES WITH CONF > 50%.

{BCE}

X	Y	CONF
B	{CE}	2/4 / 3/4 = 2/3
C	{BE}	2/4 / 3/4 = 2/3
E	{BC}	2/4 / 3/4 = 2/3
{BC}	E	2/4 / 2/4 = 1
{CE}	B	2/4 / 2/4 = 1
{BE}	C	2/4 / 3/4 = 2/3

A → B
A → D
B → C
B → D
B → E
C → D
B → A
D → A
C → B
D → B
E → B
D → E

Ex. 3. 12346 , 1257 , 14 , 2345 , 34 min supp: 40%

C1:
1 3I5
2 3I5
3 3I5
4 4I5
5 2I5
~~6 1I5~~
~~7 1I5~~

L1:
1
2
3
4
5

C2
12 2I5
~~13 1I5~~
14 2I5
~~15 1I5~~
23 2I5
24 2I5
25 2I5
34 3I5
~~35 1I5~~
~~45 1I5~~

235 , 245
1 \
23, 25, ~~35~~ 24, 25, ~~45~~

(C)

Ex. 7. {A, B, C, D, E} ABCD ACDE

a) L4
ABCD - ACDB
ACDE - ACDE
> C5
ACDBE
↓
ACDB
ACDE
BCDE
≠ L4 por isso só scan 4 vezes (L1, L2, L3, L4)

(4)

b) L4 = {ABCD, ACDE}

L3 = {ABC, ABD, BCD, ACD, ADE, CDE, ACE}

L2 = {AB, AC, AD, BC, BD, AE, CD, CE}

C3 = {ABC, ABD, ABE, ACD, ACE, ADE, BCD, CDE; ...}

C3 ≠ L3 → only for all the subsets (with 3 items) of each pattern.

Ex. 11. L4 = {ABCD, ABCE, ABDE, ACDE}

Ex. 12 {ABCD, ABCE}
↓
ABCDE

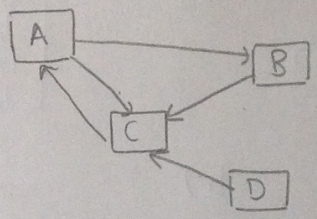
4x

Ex. 13

SOCIAL NETWORK ANALYSIS

Ex. 1.

a) distância DB → 3



b)

	IN-DEGREE	OUT-DEGREE	DEGREE
A	1	2	3
B	1	1	2
C	3	1	4
D	0	1	1

c) DEGREE CENTRALITY

$DC(A) = \frac{2}{3}$ $DC(C) = \frac{1}{3}$

$DC(B) = \frac{1}{3}$ $DC(D) = \frac{1}{3}$

CLOSENESS CENTRALITY

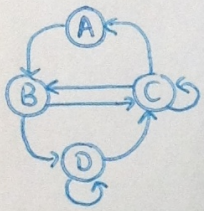
$CC(A) = \frac{3}{1+1+\infty} = \emptyset$

$CC(B) = 0$

$CC(C) = 0$

$CC(D) = \frac{3}{1+2+3} = \frac{3}{6} = \frac{1}{2}$

Ex. 2.



a) INITIALIZATION FOR PAGERANK SCORES

$$\begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$$

b) TRANSITION PROBABILITY MATRIX

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix} \end{array}$$

c) PAGERANKS

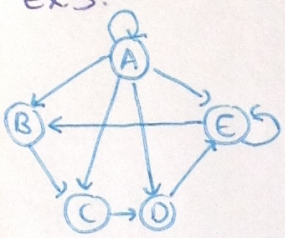
$PR(A) = 1 - 0,8 + 0,8 \times \left(\frac{1/4}{3}\right) =$

$PR(B) = 1 - 0,8 + 0,8 \times \left(\frac{1/4}{1} + \frac{1/4}{3}\right)$

$PR(C) = 1 - 0,8 + 0,8 \times \left(\frac{1/4}{3} + \frac{1/4}{2} + \frac{1/4}{2}\right)$

$PR(D) = 1 - 0,8 + 0,8 \times \left(\frac{1/4}{2} + \frac{1/4}{2}\right)$

Ex3.



a)

$$\begin{bmatrix} 1/5 \\ 1/5 \\ 1/5 \\ 1/5 \\ 1/5 \end{bmatrix}$$

b)

	A	B	C	D	E
A	1/5	1/5	1/5	1/5	1/5
B	0	0	1	0	0
C	0	0	0	1	0
D	0	0	0	0	1
E	0	1/2	0	0	1/2

c)

$$PR(A) = 1 - 0,7 + 0,7 \times \left(\frac{1/5}{5} \right)$$

$$PR(B) = 1 - 0,7 + 0,7 \times \left(\frac{1/5}{5} + \frac{1/5}{2} \right)$$

$$PR(C) = 1 - 0,7 + 0,7 \times \left(\frac{1/5}{1} + \frac{1/5}{5} \right)$$

$$PR(D) = 1 - 0,7 + 0,7 \times \left(\frac{1/5}{1} + \frac{1/5}{5} \right)$$

$$PR(E) = 1 - 0,7 + 0,7 \times \left(\frac{1/5}{1} + \frac{1/5}{2} + \frac{1/5}{5} \right)$$

Exercises - Part II

Feature selection

1.1 **Spearman correlation** (between y1 and y4) → $1 - \frac{6 \sum D^2}{n(n^2-1)} = 1 - \frac{6 \cdot (4+4+0,25+0,25)}{4(16-1)}$

y1	y4	Rank1	Rank2	D	D^2
-0,4	2	1	3	-2	4
0,7	1	3,5	1,5	2	4
0,7	3	3,5	4	-0,5	0,25
-0,3	1	2	1,5	0,5	0,25

$= 1 - \frac{6 \cdot 8,5}{60} =$

$= 1 - 0,85 = 0,15$

$\frac{3+4}{2} = 3,5$

$\frac{1+2}{2} = 1,5$

$\bar{y1} = 0,175$

$\bar{y3} = -0,325$

Pearson correlation (between y1 and y3) $= \frac{cov(y1,y3)}{\sqrt{var(y1) \cdot var(y3)}} = \frac{0,2317}{\sqrt{0,608 \times 0,478}} = 0,43$

$cov(y1,y3) = \frac{\sum (xi - \bar{x}) \cdot (yi - \bar{y})}{n-1} =$

$= \frac{(-0,4-0,175) \cdot (0,1+0,325) + (0,7-0,175) \cdot (-0,3+0,325) +}{3}$

$\frac{(0,7-0,175) \cdot (-0,1+0,325) + (-0,3-0,175) \cdot (-1+0,325)}{3}$

$= \frac{0,244 + 0,013 + 0,178 + 0,32}{3} = 0,2317$

$var(y1) = \sqrt{\frac{\sum (xi - \bar{x})^2}{n-1}} = \sqrt{\frac{0,33 + 0,276 + 0,276 + 0,226}{3}} = 0,608$

$var(y3) = \sqrt{\frac{0,18 + 0,000625 + 0,05 + 0,456}{3}} = 0,478$

1.3. Discretization

Equal-depth / frequency

(y3) → <0.1, -0,3 ; -0,1 ; -1>

y3: <1, 0, 1, 0>

[-1 ; -0,3] (0) [-0,1 ; 0,1] (1)

y2: <1 1 0 1>

Equal-width / range

$w = \frac{max - min}{2} = \frac{0,1+1}{2} = 0,55$

(y3) → <0,1 ; -0,3 ; -0,1 ; -1>

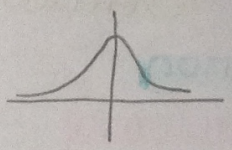
[-1 ; -0,45[(0) [-0,45 ; 0,1] (1)

<1; 1, 1, 0>

y2: <1, 1, 0, 1>

GAUSSIAN CUT-OFF POINTS

$\bar{y}_3 = -0,325$ $\quad \sigma_{y3} = 0,4145$



⚠ já estão normalizados

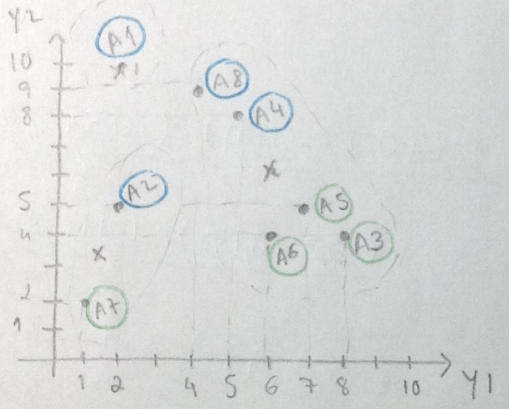
$< 0 \rightarrow 0$ $\quad <1,0,0,0>$

$> 0 \rightarrow 1$

Reduce overfitting in decision trees ← Pruning
- enforce a minimum nr of samples in leaf nodes
- enforce a maximum depth for the tree

CLUSTERING

1.



1.1. A1 , A4 , A7 → seeds ↳ centroids iniciais

↓ ↓ ↓

K1 K2 K3

- distância de cada um a A1, A4 e A7

↓

(K1) A1

(K2) A3, A5, A6, A8, A4

(K3) A2, A7

↓

K-means

1st epoch

K1 = A1

K2 = (6, 6)

K3 = (1.5, 3.5)

→ (K1): A1, A8

(K2): A4, A5, A6, A3

(K3): A2, A7

→ K1 = (3, 9.5) ✓

K2 = (6.5, 5.25)

K-median

K1 = (2,10)

K2 = (6,5) ✓

K3 = (1.5,3.5)

K-medoids

↓

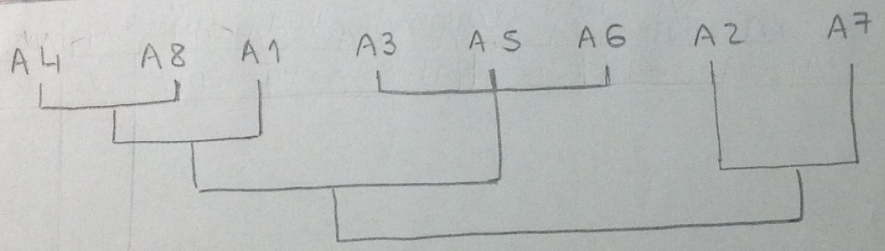
ver o ponto que está + ao centro

1.2.

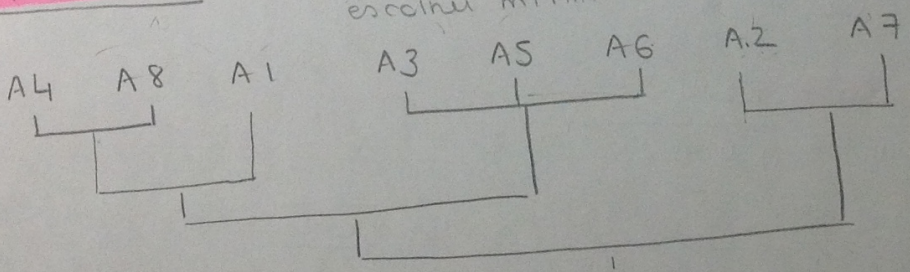
	A1	A2	A3	A4	A5	A6	A7	A8
A1	0	5	6	$\sqrt{13}$	$\sqrt{50}$	$\sqrt{52}$	$\sqrt{65}$	$\sqrt{5}$
A2		0	$\sqrt{37}$	$\sqrt{18}$	5	$\sqrt{17}$	$\sqrt{10}$	$\sqrt{20}$
A3			0	5	$\sqrt{2}$	$\sqrt{2}$	$\sqrt{53}$	$\sqrt{41}$
A4				0	$\sqrt{13}$	$\sqrt{17}$	$\sqrt{52}$	$\sqrt{2}$
A5					0	$\sqrt{2}$	$\sqrt{45}$	5
A6						0	$\sqrt{29}$	$\sqrt{29}$
A7							0	$\sqrt{58}$
A8								0

AGGLOMERATIVE: initially every point in a cluster of its own and we merge cluster until we end-up with one unique cluster containing all points

SINGLE LINK: ver as distâncias entre clusters e escolher a mínima



COMPLETE LINK: escolher máx de distância entre cluster e depois escolhe mínima dos máximos



AVERAGE LINK:

CENTROID LINK K-means

↳ calcular centro (ponto médio) de cada cluster, para distância entre cada um desses pontos e ligar o menor

1.3.

(1) PURITY (C3) $= \frac{1}{2} \cdot 1 = \frac{1}{2}$ ✓

(4) PURITY (clustering) $= \frac{1}{8}(3+3+1) = \frac{7}{8}$ ✓

(3) RI $= \frac{TP+TN}{All} = \frac{15+6}{15+6+1+6}$

TN: {A1,A3}, {A1,A5}, {A1 A6} {A4 A3} {A4 A5} {A4 A6} {A2 A3} {A2 A1} {A2 A6}
{A2,A3} {A2,A5} {A2 A6} {A7, A1} {A7 A4} {A7 A6} → 15

≠ classe em ≠ cluster

TP = {A1 A4} {A1 A6} {A4 A2} {A3 A5} {A3 A6} {A5 A6} → 6

= classe no mesmo cluster

FP = {A2 A7} → 1

estão no msm cluster e não deviam

FN = {A2A1} {A2 A4} {A2A3} {A7, A3} {A7 A5, A7 A6} → 6

estão separados e deviam estar juntos

2.

	y1	y2	cluster
x1	2	2	C1
x2	2	4	C1
x3	4	3	C2
x4	4	5	C2

(1) SILHOUETTE INDEX x2

EUCLIDEAN $a = d(x2,x1) = \sqrt{0+2^2} = 2$

$d(x2,x3) = \sqrt{2^2+1^2} = \sqrt{5}$

$d(x2,x4) = \sqrt{2^2+1^2} = \sqrt{5}$ $\Big\rangle$ $b = \frac{\sqrt{5}+\sqrt{5}}{2} = \sqrt{5} > 2$

$b > a$ $\quad s.i = \frac{b-a}{b} = \frac{\sqrt{5}-2}{\sqrt{5}} = 0,1$

CHEBYSHEV $a = d(x2,x1) = \max(|2-2|, |4-2|) = 2$

$d(x2,x3) = \max(|4-2|, |3-4|) = 2$

$d(x4,x2) = \max(|4-2|, |5-4|) = 2$ $\Big\rangle$ $b = \frac{2+2}{2} = 2$

$a = b \rightarrow$ silhouette = 0 (equally distant to C1 and C2

(3) Average Silhouette of cluster C2

Silhouette x3

$a = d(x3, x4) = |4-4| + |5-3| = 0 + 2 = 2$

$d(x3, x1) = |4-2| + |3-2| = 2+1 = 3$

$d(x3, x2) = |4-2| + |3-4| = 2+1 = 3$ $\Big\rangle b = \frac{3+3}{2} = 3$

$b > a$

$s.i(x3) = \frac{b-a}{b} = \frac{3-2}{3} = \frac{1}{3}$

Silhouette x4

$a = d(x3, x4) = 2$

$d(x4, x1) = |4-2| + |5-2| = 2+3 = 5$

$d(x4, x2) = |4-2| + |5-4| = 2+1 = 3$ $\Big\rangle b = \frac{5+3}{2} = 4$

$b > a$

$s.i(x4) = \frac{4-2}{4} = \frac{2}{4} = \frac{1}{2}$

average silhouette $(C2) = \frac{\frac{1}{3} + \frac{1}{2}}{2} = 0,417$ //

2.2. (Euclidean)

Cohesion Error (SSE) $= (2-2)^2 + (2-2)^2 + (2-3)^2 + (4-3)^2 + (4-4)^2 + (4-4)^2 + (3-4)^2 + (5-4)^2$

$= 1 + 1 + 1 + 1 = 4$

$\bar{C}_1 = (2, 3)$

$\bar{C}_2 = (4, 4)$

Separation Error (BSS) $= 2 \cdot 1,25 + 2 \times 1,25 = 5$

Ponto médio entre clusters: $\frac{\bar{C}_1 + \bar{C}_2}{2} = (3, 3.5)$

$d(C1, PM) = \sqrt{1^2 + (0,5)^2} = \sqrt{1,25}$

$d(C2, PM) = \sqrt{1^2 + (0,5)^2} = \sqrt{1,25}$

Total Error $= SSE + BSS = 4 + 5 = 9$

→

PCA

3.

$20+10+5+4+3+2+1 = 45$

$\frac{45}{45+x} = 0,9 \iff \frac{45}{0,9} = 45+x \iff x = 5$

$7+5 = \underline{12}$ attributes

4.

$$S = \begin{bmatrix} 91.43 & 171.92 & 297.99 \\ 171.92 & 373.92 & 545.21 \\ 297.99 & 545.21 & 1297.26 \end{bmatrix}$$

$S\vec{u_1} = \lambda_1 \vec{u_1} \iff (S - \lambda_1 I).\vec{u_1} = 0$

$$\begin{bmatrix} 91.43 & 171.92 & 297.99 \\ 171.92 & 373.92 & 545.21 \\ 297.99 & 545.21 & 1297.26 \end{bmatrix} \begin{bmatrix} 0.22 \\ 0.41 \\ 0.88 \end{bmatrix} = \begin{bmatrix} 352,833 \\ 670,91 \\ 1430,68 \end{bmatrix}$$

$$\begin{bmatrix} 352,833 \\ 670,91 \\ 1430,68 \end{bmatrix} = \begin{bmatrix} \lambda_1 \cdot 0,22 \\ \lambda_1 \cdot 0,41 \\ \lambda_1 \cdot 0,88 \end{bmatrix} \iff \lambda_1 = 1603,7 \quad \lambda_2 = 127,65$$
$$\lambda_1 = 1636,37 \quad \lambda_2 = 129,45$$
$$\lambda_1 = 1625,77 \quad \lambda_2 = 127,85$$

$\lambda_3 = 7,543$

$$\frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} = \frac{1603,7 + 127,65}{1603,7 + 127,65 + 7,543} = 0,99$$

5.

(1) **TRANSFORM THE INPUT** (remover média)

Transformed data = Featurevector^T x Data Adjusted

$$= \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 2 & -1 & -1 \\ 0 & 1 & -1 \end{bmatrix}$$

(2) **Recover THE ORIGINAL DATA** using the most informative component

y_1 pq $\lambda_1 > \lambda_2$

$$\frac{1}{\sqrt{2}}\begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

REGRESSION

1.1.

b) MSE = $\frac{\sum (z_i - \hat{z}_i)^2}{n}$ = $\frac{4{,}813^2 + (-1{,}231)^2 + (-0{,}081)^2 + (-4{,}951)^2 + 0{,}387^2 + 0{,}163^2}{6}$ (residuos)

= 8,24

RMSE = $\sqrt{8{,}24}$ =

MAE = $\frac{1}{n} \sum |z_i - \hat{z}_i|$

MAPE = $\frac{1}{n} \sum \left| \frac{z_i - \hat{z}_i}{z_i} \right|$

1.3. $d(x_i, x_j) = |10(a_{i1} - a_{j1}) + (a_{i2} + a_{j2})/10|$ || similaridades distancia weights

(3) $s_{11} = 1 - \frac{d}{d_{max}} = 1 - 0 = 1$

$s_{12} = 1 - \frac{d(x_1, x_2)}{100} = 1 - \frac{74}{100} = 0{,}26$ ...

$s = s_{11} + s_{12} + s_{13} + s_{14} + s_{15} + s_{16}$

(4) distancia weights por x_1

$w_{12} = \frac{s_{12}}{s}$ $w_{13} = \frac{s_{13}}{s}$

(5) $k = 2$

$x_1 \times 5$ $x_1 \times 6$

$w_{15} = 0{,}39$ $w_{16} = 0{,}61$

$0{,}39 \times 13{,}75 + 0{,}61 \times 18{,}11$

BICLUSTERING

4.

SEQUENTIAL PATTERN MINING

1.

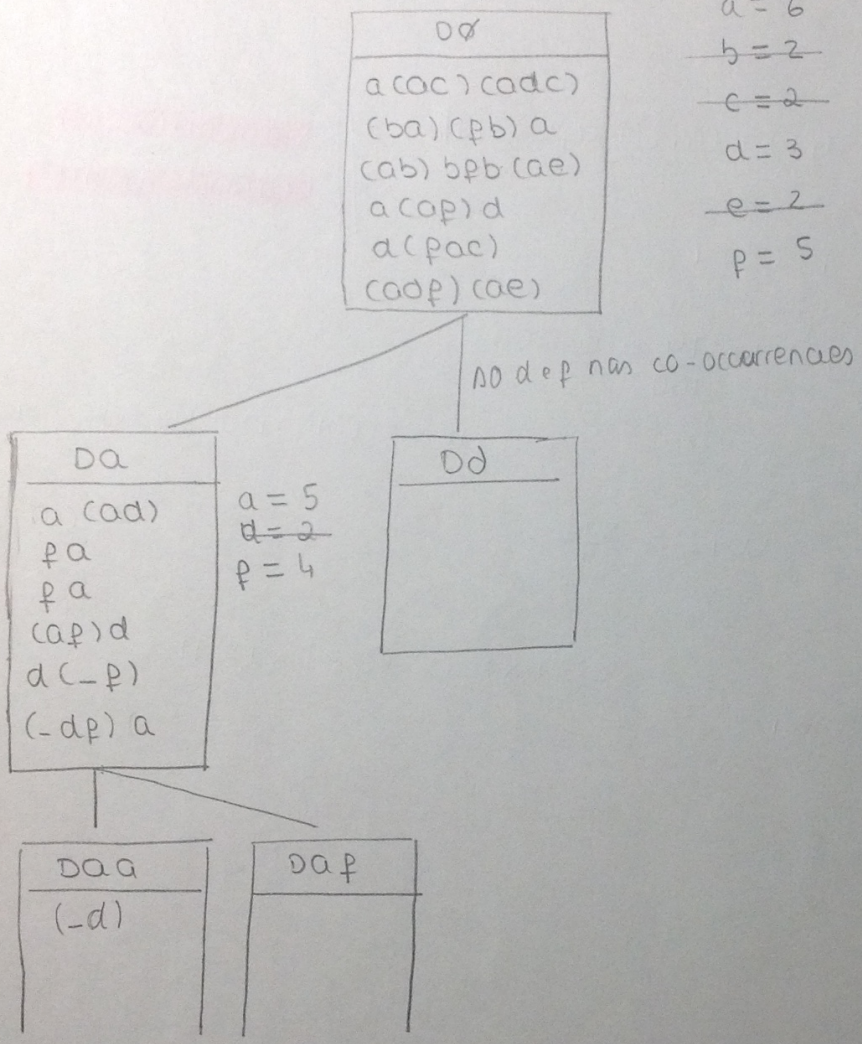
(1) $\{1,2,3,4\} \longrightarrow$ 4 items ✓

maximal seq. pat : not a subsequence of a seq. pattern

closed seq. pat : " " " " " " " with same support

2. PREFIXSPAN

min abs supp = 3



Time series forecasting

1. (1) residuals increasing → multiplicative model

(2) Irregular component:

- baseline
/ seasonal index

$< -1,42 ; 0 ; 2,67, \dots >$

2.

$<1, 6, 10, 13, 16, 19, 21, 23>$

(1) $\hat{x}(10) = 23$

(2) one differencing

$<1, 6, 10, 13, 16, 19, 21, 23>$

$<5, 4, 3, 3, 3, 2, 2>$

$\hat{x}(9) = 23 + 2 = 25 \qquad \hat{x}(10) = 27$

(3) $\hat{x}(t) = 2x(t-2) - 1,2x(t-1)$

$\hat{x}(9) = 2x(7) - 1,2x(8) = 2 \times 2 - 1,2 \times 2 = 4 - 2,4 = 1,6$

$\hat{x}(10) = 2x(8) - 1,2x(9) = 2 \times 2 - 1,2 \times 1,6 = 2,08$

$\hat{x}(10) = 23 + 1,6 + 2,08 = 26,68 = 26,7$

(5) $\hat{x}(t) = \underbrace{2x(t-2) - 1,2x(t-1)}_{p} + \varepsilon(t-1)$ (9) $\qquad d = \#$ differings

3. Mean Forecasting Error $= \frac{\sum x - \hat{x}}{n} = \frac{2+3+1+2+3+3+3+1+2+4+1+2}{12} =$

$= \frac{7}{12} = 0,583$

Mean absolute deviation $= \frac{\sum |x - \hat{x}|}{n} = \frac{2+3+1+2+3+3+3+1+2+4+1+2}{7} =$

$= \frac{27}{12} = 2,25$

→D

Time series representations

1.

1.1. (2) PAA of x1

<1,4,2,2,4,1>

↓ normalizar (−média / σ)

<−1,07 ; 1,34 ; −0,26 ; −0,26 ; 1,34 , −1,07>

6 segmentos → 5 segmentos

$\frac{6}{5} = 1,2$

$\frac{-1,07 \times 1 + 1,34 \times 0,2}{1,2} = -0,67$

$\frac{1,34 \times 0,8 + (-0,26) \times 0,4}{1,2} = 0,80$

$\frac{(-0,26) \times 0,6 + (-0,26) \times 0,6}{1,2} = -0,27$

$\frac{(-0,26) \times 0,4 + 1,34 \times 0,8}{1,2} = 0,80$

$\frac{1,34 \times 0,2 + (-1,07) \times 0,8}{1,2} = -0,67$

→ <−0,67 ; 0,8 ; −0,27 ; 0,8 ; −0,67>

(4)

<0,1,2,3> <1,3,4,2>

CHEBYSHEV : max(|0−1|, |1−3|, |2−4|, |3−2|) = 2

(5) X2 Hamming

<1,3,4,2>

1.2.

(2) <abgba> → abc motif with support = 2

(3)

aa	ab
ba	bb

0	2/4
2/4	0

(4) <ababa>

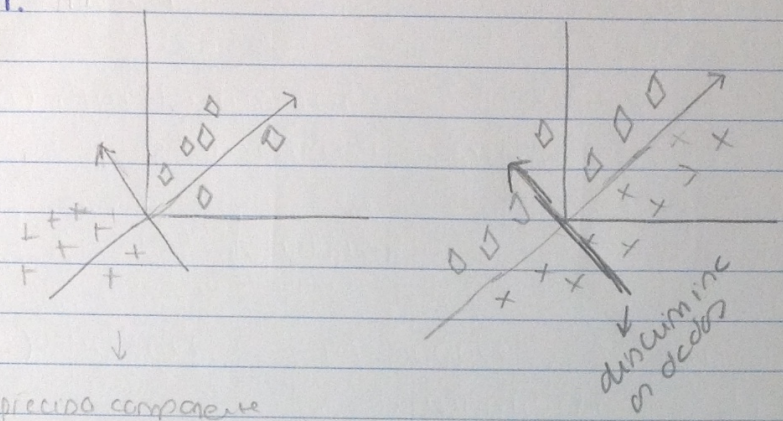
aba → 2/3 bab → 1/3

aaa 0
aab 0
aba 2
abb 0
baa 0
bab 1
bba 0
bbb 0

PRINCIPAL COMPONENT ANALYSIS

> eigenvector → > variance

1.



↓

precipa componente

↓ > eigenvalue

(ii) one of the principal components can accurately discriminate the class.

2. (3)

3. TOP-7 eigenvalues explain 90%.

$\lambda 1 = 20$

$\lambda 2 = 10$

$\lambda 3 = 5$

$\lambda 4 = 4$

$\lambda 5 = 3$

$\lambda 6 = 2$

$\lambda 7 = 1$

$20^2 + 10^2 + 5^2 + 4^2 + 3^2 + 2^2 + 1^2 =$

$= 400 + 100 + 25 + 16 + 9 + 4 + 1 =$

$= 555$ ✗

$$\frac{45}{45+x} = 0,9 \Leftrightarrow 45 + x = \frac{45}{0,9} \Leftrightarrow$$

$$\Leftrightarrow 45 + x = 50 \Leftrightarrow x = 5$$

↓

7 + 5 attributes

↓

12 attributes

4.

covariance matrix $S = \begin{bmatrix} 91,43 & 171,92 & 297,99 \\ & 373,92 & 545,21 \\ & & 1297,26 \end{bmatrix}$

EIGENVECTORS:

$\vec{u_1} = \begin{bmatrix} 0,22 \\ 0,41 \\ 0,88 \end{bmatrix}$ $\vec{u_2} = \begin{bmatrix} 0,25 \\ 0,85 \\ -0,46 \end{bmatrix}$ $\vec{u_3} = \begin{bmatrix} 0,94 \\ -0,32 \\ -0,08 \end{bmatrix}$

$S\vec{u_1} = \lambda_1 \vec{u_1}$

$(S - \lambda_1 I)\vec{u_1} = 0 \Leftrightarrow$

$$\begin{bmatrix} 91,43 & 171,92 & 297,99 \\ & 373,92 & 475,21 \\ & & 1297,26 \end{bmatrix} - \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_1 \end{bmatrix} \vec{u_1} = 0$$

$$\begin{bmatrix} 91,43-\lambda_1 & 171,92 & 297,99 \\ 0 & 373,92-\lambda_1 & 475,21 \\ 0 & 0 & 1297,26-\lambda_1 \end{bmatrix} \begin{bmatrix} 0,22 \\ 0,41 \\ 0,88 \end{bmatrix} = 0$$

$$\begin{bmatrix} (91,43-\lambda_1)\times 0,22 + 171,92 \times 0,41 + 297,99 \times 0,88 \\ 0 + (373,92-\lambda_1)\times 0,41 + 475,21 \times 0,88 \\ (1297,26-\lambda_1)\times 0,88 \end{bmatrix} = 0$$

$$= \begin{bmatrix} 20,1146 - 0,22\lambda_1 + 70,4872 + 262,2312 \\ 153,3072 - 0,41\lambda_1 + 418,1848 \\ 1141,5888 - 0,88\lambda_1 \end{bmatrix} = 0$$

$$= \begin{bmatrix} 352,833 - 0,22\lambda_1 \\ 571,492 - 0,41\lambda_1 \\ 1141,5888 - 0,88\lambda_1 \end{bmatrix} = 0$$

$-0,22\lambda_1 = -352,833 \Leftrightarrow \lambda_1 = 1603,78636$

$-0,41\lambda_1 = -571,492 \Leftrightarrow \lambda_1 = 1393,88$

$-0,88\lambda_1 = -1141,5888 \Leftrightarrow \lambda_1 = 1297,26$

5.

	y_1	y_2
x_1	1	−1
x_2	0	1
x_3	−1	0

$\lambda_1 = 3 \quad \lambda_2 = 1$

$\vec{u_1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \vec{u_2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

(1) TRANSFORM THE INPUT DATA USING PCA

DataAdjusted $= \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix}$

featurevector $= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

$$\text{Transformed data} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}^T \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix} =$$

$$= \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}^T \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix} =$$

$$= \begin{bmatrix} 2/\sqrt{2} & -1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

(2) RECOVER THE ORIGINAL DATA

Data Recovered = (Feature Vector x TransformedData) + OriginalMean

Data recovered = DataAdjusted + Original mean

Data recovered

Regression

1.

$\hat{z} = \hat{\beta}_0 + \hat{\beta}_1 \cdot y_1 + \hat{\beta}_2 y_2 \rightarrow$ MULTI

a) Residuals?

4 erros $\quad \hat{z} = 2{,}341 + 1{,}616\, y_1 + 0{,}014\, y_2$

$\boxed{\hat{z} - z}$

$\hat{z}_1 = 2{,}341 + 1{,}616 \times 7 + 0{,}014 \cdot 560 = 21{,}493$

$\hat{z}_2 = 2{,}341 + 1{,}616 \times 3 + 0{,}014 \cdot 220 =$.

$21{,}493 - 16{,}68 = 4{,}813$

b) $MSE(\hat{z}, z) = \frac{1}{n} \sum_{i=1}^{\hat{}} (z_i - \hat{z}_i)^2 =$

$= \frac{1}{6} \cdot \left[4{,}813^2 + (-1{,}231)^2 + (-0{,}081)^2 + (-4{,}955)^2 + 0{,}387^2 + 0{,}163^2 \right] = 8{,}24$

$RMSE(\hat{z}, z) = \sqrt{\frac{1}{n} \hat{\sum} (z_i - \hat{z}_i)^2} = \sqrt{8{,}24} = 2{,}8698$

$MAE(\hat{z}, z) = \frac{1}{n} \sum |z_i - \hat{z}| = \frac{1}{6} \left(4{,}813 + 1{,}231 + 0{,}081 + 4{,}955 + 0{,}387 + 0{,}163 \right)$

$= 1{,}9383$

$MAPE(\hat{z}, z) = \frac{1}{n} \sum \left| \frac{z_i - \hat{z}_i}{z_i} \right| = \frac{1}{6} \left(\frac{4{,}813}{16{,}68} + \frac{1{,}231}{11{,}5} + \right.$

$\downarrow$

$\times \quad \left. \frac{0{,}081}{12{,}03} + \frac{4{,}955}{14{,}88} + \frac{0{,}387}{13{,}75} + \frac{0{,}163}{18{,}11} \right) \simeq 0{,}1287$

(3)

1.2. $y = mx + b \rightarrow$ univariate

(1)

1.3. $d(x_i, x_j) = |10(a_{i1} - a_{j1}) + (a_{i2} - a_{j2})/10|$

$d(x_1, x_2) = 74$
$d(x_1, x_3) = 62$
$d(x_1, x_4) = 78$
$d(x_1, x_5) = 51$
$d(x_1, x_6) = 23$
$d(x_1, x_4) = 4$
$d(x_1, x_3) = 12$

(1) ver distância menor em cada x e onde z é o $\hat{z}$

(2) ver dois pontos mais próximos ver valores de z, fazer média e é o $\hat{z_i}$

(3) ← similaridade

$s = 1 - \frac{d}{d_{max}}$

$s_{11} = 1 - 0 = 1$

$s_{12} = 1 - \frac{74}{100} = 1 - 0{,}74 = 0{,}26$

...

(4)

distância weight (ou w parte) → é soma das similaridades duma ponto das vizinhas escolhidas

$0{,}26 + 0{,}38 + 0{,}22 + 0{,}45 + 0{,}77$

é $\frac{\text{similaridade}}{\text{soma (total)}}$

$w_{12} = \frac{s_{12}}{\text{soma singularidades}}$

(5)

$\hat{z_1} = 13{,}75 \times 0{,}39 + 18{,}11 \times 0{,}61 = 16{,}4$

$k = 2$

$s_{15} + s_{16} = 0{,}49 + 0{,}77 = 1{,}26$

$w_{15} = \frac{0{,}49}{1{,}26} = 0{,}39$

$w_{16} = \frac{0{,}77}{1{,}26} = 0{,}61$

(6)

2.

2.1. $y = \beta_0 + y_1\beta_1 + y_2\beta_2$ (3)

$2 \times 0{,}8 \quad + 0$

↓

without regard to y_2

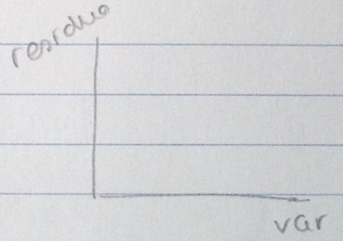
↓

varia!!

2.2.

least square error. – (3)

3.



– se o modelo está biased

– se erros são muito altos

Ideal: erro

↓

Tendência: MAU

↓

p̄ aqueles valores o estimador tem maus resultados

↓

subestima sempre o valor real

(4)

x_5 x_6

2 + prox

$\frac{13{,}75 + 18{,}11}{2} =$

BICLUSTERING : relação entre variáveis (encontra)

variáveis →
instâncias ↓

constant
additive
multiplicative
order-preserving

1. I: instâncias
   J: features

(1) ORDER-PRESERVING:

	y2	y3	y5
x1	2	5	4
x2	1	4	2
x3	2	5	4

vê como variam os features $y_2 \leq y_5 \leq y_4$

(2) CONSTANT ✓

	y2	y3	y5
x1	2	5	4
x3	2	5	4

(3) ADDITIVE : $\bar{p}$ todas as colunas a soma é a mesma

φ_B = 2 1 4 2

	y1	y2	y3	y4	
x1	3	2	5	3	$\varphi_B + 1$
x2	2	1	4	2	$\leftarrow \varphi_B + 0$

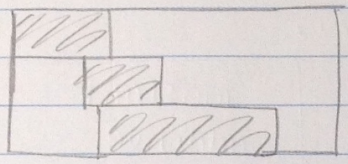
(4) MULTIPLICATIVE

1 4 2

	y2	y3	y5	
x1	2	5	4	y ✗
x2	1	4	2	$\leftarrow \varphi_B \times 1$ (4)
x3	2	5	4	$\leftarrow \varphi_B \times 2$ ✗

tapaiporrusal novasrotasTAP

2.



(3) (4)
(5)
(7)

3. (3) false negative!!

4. $\bar{A}$ ← amplitude of the range of values in a matrix A

Given A, coerence strength is a range $\delta \in (0,1)$ such that

$a_{ij} = c_j + \gamma_i + \eta_{ij}$ (or other) and (noise)

$\eta_{ij} \in [-\delta/2, \delta/2]$

(1) constante = 2,1 pq noise = η = ∅ NOT ERROR → noise

(2) 3,2 + 0,1 = 3,3 < 3,4 $\eta = (-0,1; 0,1)$

(3)

1,2 x2
1,2 x4
1,2 x7
0 x8
1,2 x9

	y3	y6	y8	y9
x8	0	1	3,1	0,5
x2	1,2	2,2	3,3	1,7
x4				
x7				
x9				

$p \pm 0,2$

← $\gamma_8 = \emptyset$
$\gamma_2 = 1,2$
$\gamma_4 = 1,2$
$\gamma_7 = 1,2$
$\gamma_9 = 1,2$

(4) Quality: amount of noisy elements + type of noise → $[\eta_{ij} - \delta/2, \eta_{ij} + \delta/2]$

ni: nº of elements with minimum coherence strength

$\frac{16-2}{16} = 0,875$

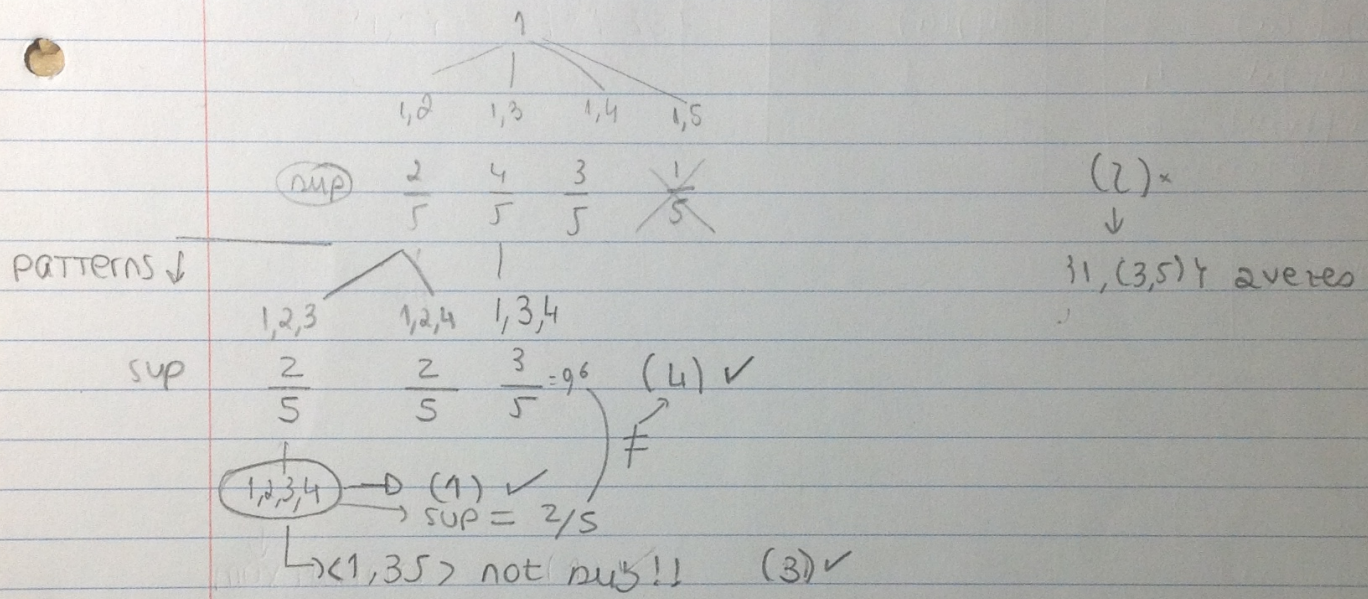
SEQUENTIAL PATTERN MINING

1.

SID	Sequence
1	<{1,5}{2}{3}{4}>
2	<{1}{3}{4}{3,5}>
3	<{1}{2}{3}{4}>
4	<{1}{3,5}{5}>
5	<{4}{5}>

GSP //

minsup = 0,4 — 3 items per pattern



(2) ×
↓
{1,(3,5)} 2 veces

2.

SID	Sequence
1	<a(ac)(acd)>
2	<(ba)(fb)a>
3	<(ab)bfb(ae)>
4	<a(af)a>
5	<d(fac)>
6	<(adf)(ae)>

minimum absolute support → 3

PREFIXSPAN

Dσ	
1	a(ac)(adc)
2	(ba)(fb)a
3	(ab)bfb(ae)
4	a(af)d
5	d(fac)
6	(aaf)(ae)

a=6
~~b=2~~
~~c=2~~
d=3
~~e=2~~
f=5

co-occurring é só uma posição!

Da		
	(a)(ad)	1
	(f)a	2
	f(a)	3
	(af)d	4
	(a)	6

a=5
d=1
f=3

Dd			
	(f(fa)	5	2=a
	(f)(a)	6	2=f

Df		
a	2	
(a)	3	a=3
d	4	~~d=1~~
(a)	6	

Daa	
(a)	1

Da
(a)(ad)
(f)a
f(a)
(af)d
(_f)
(_df)(a)

TIME SERIES FORECASTING

1. <2, 3, 7, 4, 4, 6, 3, 4, 8> Time series
   4 item trend

   trend component was subtracted → ADDITIVE

ADDITIVE MODEL

DATA = seasonal effect + Trend + Cyclical + Residual

MULTIPLICATIVE MODEL

DATA = (seasonal effect) x Trend x Cyclical x Residual

seasonal indexes : <0.7, 0.8, 1.5>

RESIDUALS

(1) Residuals increase along time suggesting that the trend should be divided instead of subtracted

(2) subtract baseline and remove seasonal component by divide values by the seasonal indexes

(3) Residuals increase along time

$\hat{x}(t) = -t + t^2$

$2 + (-1) + 1^2 = 2$

$3 + (-2) + 2^2 = 5$

$7 + (-3) + 3^2 = 13$ ... (4) ✓

(2)

$\frac{2-3}{0,7} =$

$\frac{3-3}{0,8} =$

$\frac{7-3}{1,5} =$ ✓

$\frac{4-3}{0,7} =$

⋮

2. Time series: <1, 6, 10, 13, 16, 19, 21, 23>

naive forecast → usa último valor da mund e
pat linha reta
(não sime ncda)

(1) ✓

DIFFERENCING → diferença entre cada ponto
(6-1, 10-6, 13-10)...

two time points ahead: 10-1, 13-6, 16-10...

23-19 = (4) $\hat{x}(10) = 23 + 4 = 27$

(2) ✓ percorre ponto a ponto 1 diferencing
<5, 4, 3, 3, 3, 2, 2>

$\hat{x}(9) = 23 + 2 = 25$ 25-23=2

$\hat{x}(10) = 25 + 2 = 27$

naive forecast escolhe último

(3)

$\hat{x}(9) = 2 \times x^{DIFF}(7) - 1,2\, x^{DIFF}(8)$

$= 2 \times 2 - 1,2 \times 2 = 1,6$

$\hat{x}(10) = 2 \times x^{DIFF}(8) - 1,2 \times x^{DIFF}(9)$

$= 2 \times 2 - 1,2 \times 1,6 = 2,08$

$23 + 1,6 + 2,08 = 26,68$

(4) one differencing

1	2	3	4	5	6	7	8	
1	6	10	13	16	19	21	23	← (8)
	5	4	3	3	3	2	2	
							ε(8)	

$\hat{x}(9) = 0,8\, x(7) +$

(5) p: nº tempos p não usamos
d: # differings
q: $\varepsilon(t-q)$

ARIMA
p: AR order
d: integration order - # differencings to stationarize time series
q: MA order

3.

$\langle 2, 3, -1, -2, 3, 3, -3, -1, 2, 4, -1, -2 \rangle$ Resíduos

(1) mean forecasting error

$$MFE = \frac{\sum x_t - \hat{x}_t}{n} = \frac{7}{12} = 0{,}583$$ ✓

(2) mean absolute deviation

$$MAD = \frac{\sum |x_t - \hat{x}_t|}{n} = \frac{27}{12} = 2{,}25$$ ✓

(3)

$12/3 = 4$ ↓ grupos 4

++ −− → BIAS (a cada 4) ✓

(4) X

Time series representation

1.

$x_1 = \langle 1,4,2,2,4,1 \rangle$ $\quad x_2 = \langle 1,3,4,2 \rangle$

1.1.

(1) $\bar{x_2} = \frac{1+3+4+2}{4} = \frac{10}{4} = 2,5$ $\quad \sigma_{x2} \simeq 1,12$

$x = x_1 + x_2$ $\quad \sigma_x = 1,2$

$\bar{x} = \frac{1+4+2+2+4+1+1+3+4+2}{10} = \frac{24}{10} = 2,4$

$x_2' = \langle -1,34, \ldots \rangle$

$x_2'' = \langle -1,17, \ldots \rangle$

ponto-média / σ

(2)

PAA representation $\quad \bar{x_1} = \frac{14}{6} \simeq 2,33$

↓

Piecewise aggregate approximation $\quad \sigma_1 = 1,247$

$x_1 = \langle 1,4,2,2,4,1 \rangle$ $\quad \frac{6}{5} = 1,2$

↓ normalizar

$\langle -1,07 ; 1,34 ; -0,26 ; -0,26 ; 1,34 ; -1,07 \rangle$

↓ 5 segmentos

$\frac{-1,07 + 1,34 \times 0,2}{1,2} \simeq -0,67$ $\quad \frac{1,34 \times 0,8 + (-0,26) \times 0,4}{1,2}$

$5 \times (-1,07) + 1 \times 1,34 = -4,01$

$4 \times (1,34) + 2 \times (-0,26) =$

$3 \times (-0,26) \times 2 =$

$2 \times (-0,26) + 4 \times 1,34 =$

$1 \times (1,34) + 5 \times (-1,07) =$

$\langle -\frac{4,01}{6}, \frac{4,84}{6}, \frac{-1,56}{6}, \frac{4,84}{6}, \frac{-4,01}{6} \rangle$

$\langle -0,67 , 0,80 , -0,27 ; 0,80 , -0,67 \rangle$ ✓

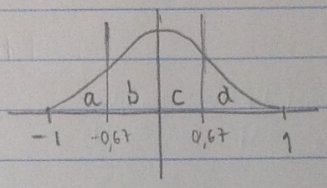
(3) SAX Representation : x2

$\bar{x}_2 = 2,5 \quad \sigma_2 = 1,12$

4 segmentos

normalizado: <-1,34 ; 0,45 ; 1,34 ; -0,45>

DISTRIBUIÇÃO normal:



<a, c, d, b> ✓

(4) BIJECTIVE MAPPING

<0,2,3,1> x_2 = <1,3,4,2> ✗

CHEBYSHEV: (1)

(5) HAMMING DISTANCE

↓

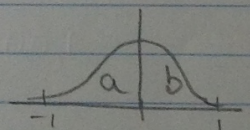
normalized by the number of segments

1.2.

(1) CODEBOOK SYMBOLIZATION

(2) SAX representation

<-0,67 ; 0,80 ; -0,27 ; 0,80 ; -0,67>
a b a b a



aba → MOTIF: conjunto dessas letras comum ✓

support 2 → aparece 2x ✓

(3) SECOND-ORDER BITMAP

2 combinações de letras

aa	ab
ba	bb

→

0/4	2/4
2/4	0

pq 2x ab + 2x ba

Δ normalizar

1 mark frequencies

2 normalize

(4) THIRD-ORDER BITMAP

aaa	aab	aba
abb	baa	bab
bba	bbb	

aba → 2x

$\frac{2}{3}$ → porque aba 2 vezes e bab 1 vez

(5) Lower BOUNDING PROPERTY: distances on the new space always lower than in original space
Upper: contrário

1.3

T21 → DTW distance: $\min \left\{ \sqrt{\sum_{k=1}^{K} w_k} / K \right.$

(1) $x1 = <1,4,2,2,4,1>$

$x2 = <1,3,4,2>$

computed distance = 4

x2 \ x1	1	4	2	2	4	1
2	14	4	1	1	5	4
4	13	1	5	6	3	12
3	4	1	2	3	4	8
1	1.0	9	10	11	20	20

$\gamma(i,j) = d(q_i, c_j)^2 + \min \{ \gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1) \}$

(2) MINIMUM LENGTH → 6 alignments
↳ n.º elementos

(3) ✓

2.1.

(1) Fourier Transform
→ When applying the DFT over a signal, the sampling rate needs to be high enough to adequately model high frequencies

(2) → DFT decomposes a signal in terms of its sinusoids

(3) → DFT representations offer the possibility to derive tabular data mapping from complex signals since temporal information is lost

(4) → X SDFT not preferred over DFT for stationary time series
↓
for non-stationary

(5) Different types of signals are optimally described by ≠ types of wavelets, such as Haar, morlet or Daubechies waveforms

3. DTW is prefered over DFT:
   → signals with discontinuities (abrupt changes)
   → signals with drupts and trends
   → when the length of stationary segments within a signal is unknown or varying

TEMPORAL DATA MINING

1.

	time series
x_1	<3,2,6>
x_2	<2,5,4,2>
x_3	<3,5,5>
x_4	<3,7,6,3>

1.1. K-medoids – x_1 and x_4

medoid → observation that minimizes the total distances within the cluster

DTW with square penalization

(1)

	6	17	2	6	22
	2	1	10	9	6
x_1	3	1	5	6	7
x_2		2	5	4	2

como 22 > 9 ✗

$$\frac{\sqrt{1+1+2+6+22}}{5} = 1,731$$

(2)

	6	10	2	3
	2	1	9	18
x_1	3	0	9	18
x_3		3	5	5

$$\frac{\sqrt{0+1+2+3}}{4} = 0,612$$

	3	43	16	7	8
	6	42	6	9	25
	7	26	5	14	31
x_4	3	1	5	8	7
x_2		2	5	4	2

$$\frac{\sqrt{1+5+5+6+7+8}}{6} = 0,94$$

	3	25	9	9
	6	25	5	5
	7	16	4	8
x_4	3	0	4	8
x_3		3	5	5

$$\frac{\sqrt{0+4+5+9}}{4} = 1,06$$

(3)

(4) K-means is not adequate method to be appued on time series since the mean operation can cancel waveforms in the presence of misalignments

(5) K-means algorithm fits data in hypersphere distance spaces

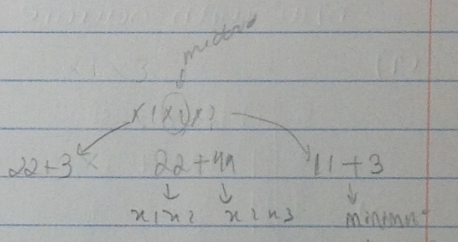
(6) EM algorithm fits data in hyperelipsoid distance spaces

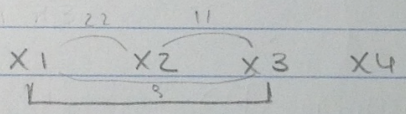
1.2.

(1) AGGLOMERATIVE + SINGLE LINK

matriz de distâncias usando DTW

	X1	X2	X3	X4
X1	0	22	(3)	11
X2	22	0	11	8
X3	3	11	0	9
X4	11	8	9	0





	X1	X2	X3	X4
X1	0	8	3	5
X2	8	0	5	6
X3	3	5	0	5
X4	5	6	5	0

↓

MATRIZ (STOR)

TEMPORAL DATA MINING

1.1.

(1) min : 3

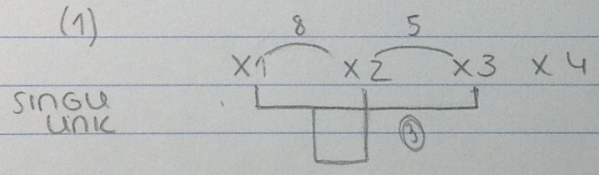
max : 8

(2)

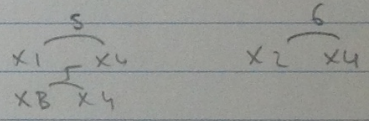
X1 — X3 X4 — X2

1.2. AGGLOMERATIVE CLUSTERING

(1)

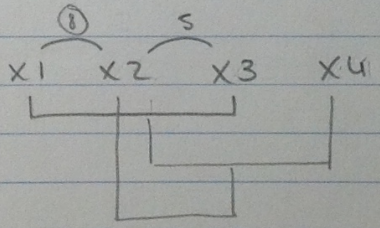


SINGLE LINK

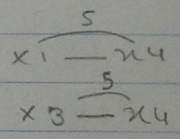


empate!

(2)



COMPLETE LINK



escolher mínimo entre os máximos

(3) SILHOUETTE SCORE

{x1, x2, x3 4}

d(x1, x2) = 8

d(x1, x3) = 3

d(x2, x3) = 5

d(x1, x4) = 5

d(x2, x4) = 6

d(x3, x4) = 5

$a_{x1} = \frac{8+3}{2} = \frac{11}{2} = 5,5$ $b_{x1} = 5$ $s = \frac{b}{a} - 1 = \frac{5}{5,5} - 1 = -0,09$

$a_{x2} = \frac{8+5}{2} = \frac{13}{2} = 6,5$ $\quad b_{x2} = 6$ $\quad s = \frac{6}{6,5} - 1 = -0,0769$

$a_{x3} = \frac{3+5}{2} = \frac{8}{2} = 4$ $\quad b_{x3} = 5$ $\quad s = 1 - \frac{4}{5} = 0,2$

$a_{x4} = 0$ $\quad b_{x4} = \frac{5+6+5}{3} = \frac{16}{3}$ $\quad s = 1$

$\frac{1 + 0,2 - 0,0769 - 0,09}{4} = 0,26$??

2.

	time series	z
x1	$\begin{pmatrix} y1 \\ y2 \end{pmatrix} = \begin{pmatrix} 3,2,2 \\ 2,3,1 \end{pmatrix}$	1,2
x2	$\begin{pmatrix} y1 \\ y2 \end{pmatrix} = \begin{pmatrix} 2,3,2,3 \\ 2,3,4,1 \end{pmatrix}$	0,9
x3	$\begin{pmatrix} y1 \\ y2 \end{pmatrix} = \begin{pmatrix} 2,2,2 \\ 4,4,1 \end{pmatrix}$	0,6

DTW to measure distances with manhattan distance

↓

loss function between the feature vectors on two time points

2.1.

(y1)

	2	1	2	1	(2)
	2	1	2	[1]	2
x1	3	[1]	[1]	2	2
y2		2	3	2	3

(y2)

	1	2	2	3	(1)
	3	1	[0]	[1]	3
x1	2	[0]	1	3	4
x2		2	3	4	1

(2) ✓

(3) DTW não parametriza !!

Regressão do KNN com time series ☺

(4)

2.2. Lazy Regression – KNN

(1) uniform weights → mean operator

(X1) – x2 ← mais próximo → $\hat{z}_1 = 0,9$

\x3

x2 – x1

\x3

$\hat{z}_2 =$?

(2) $RMSE = \sqrt{\frac{1}{n} \Sigma (z_i - \hat{z}_i)^2} =$

$= \sqrt{\frac{1}{3}[(1,2-0,9)^2 + (0,9-1,2)^2 + (0,6-0,9)^2}$

$= \sqrt{\frac{1}{3}(0,09 + 0,09 + 0,09)} = 0,3$ ✓

$MAPE = \frac{1}{3} \cdot (0,25 + 0,33 + 0,5) = 0,36$

(3) $s = 1 - \frac{d}{dmax}$

$s_{11} = 1 - 0 = 1$

$s_{12} = 1 - \frac{1}{10} =$

(4) $\hat{z}$