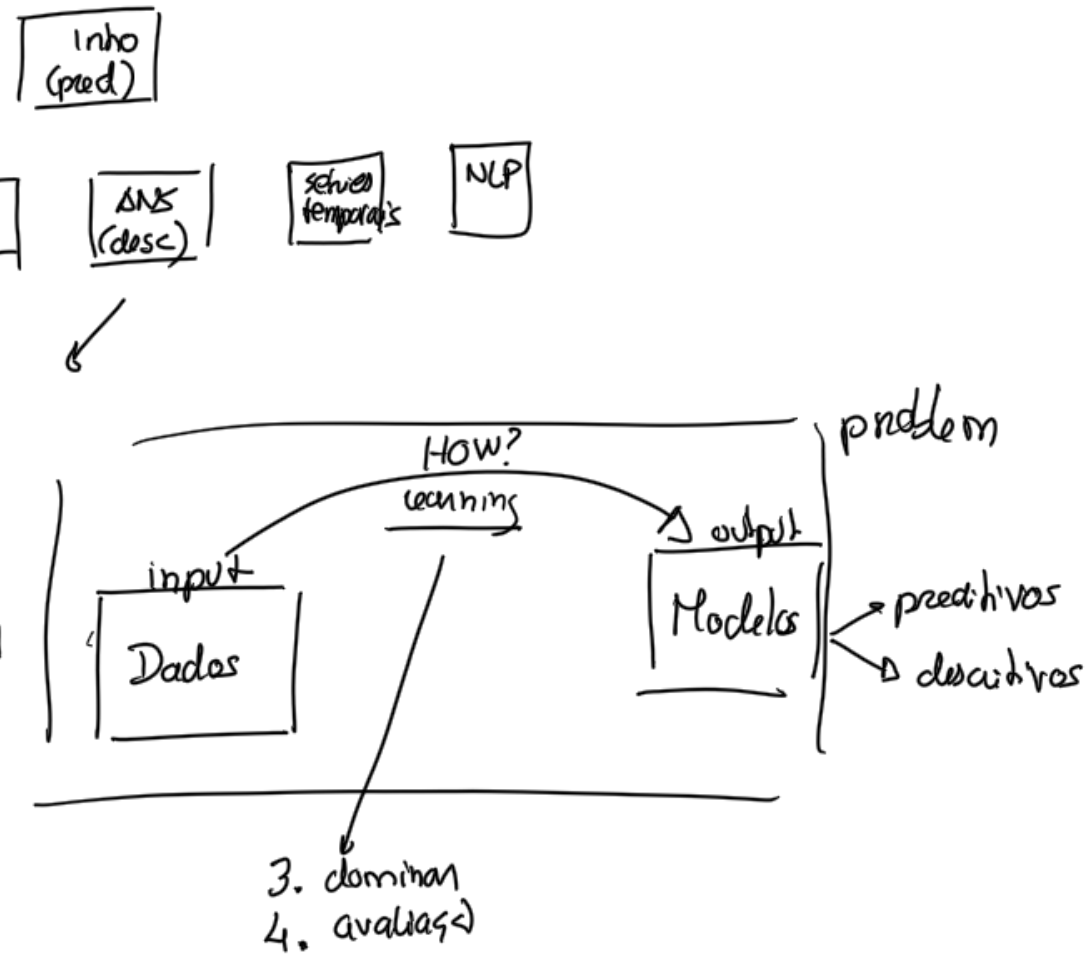
6 1) organização → 2) aprend. não sup
→ 3) estatística descritiva
→ 4) clustering

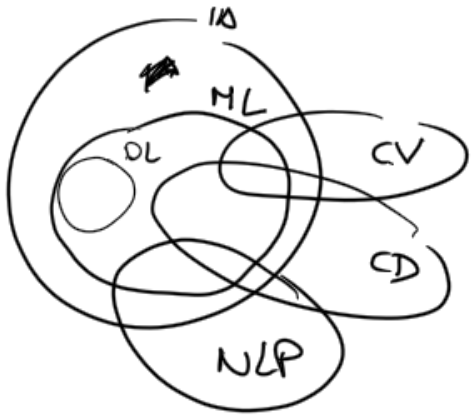
DASH



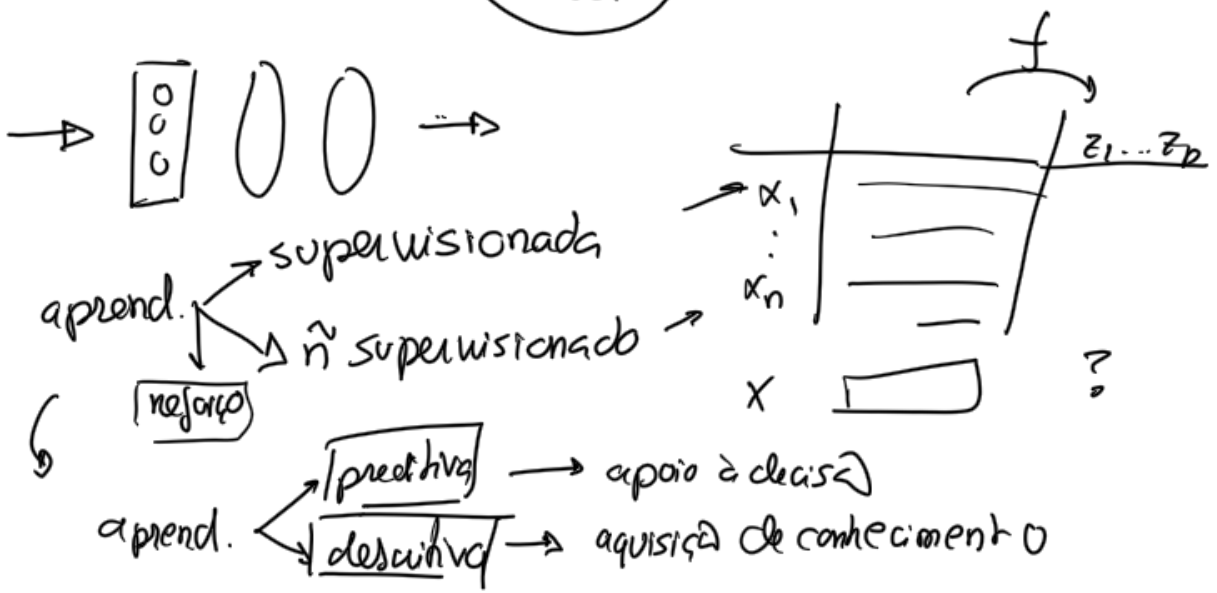
Introdução

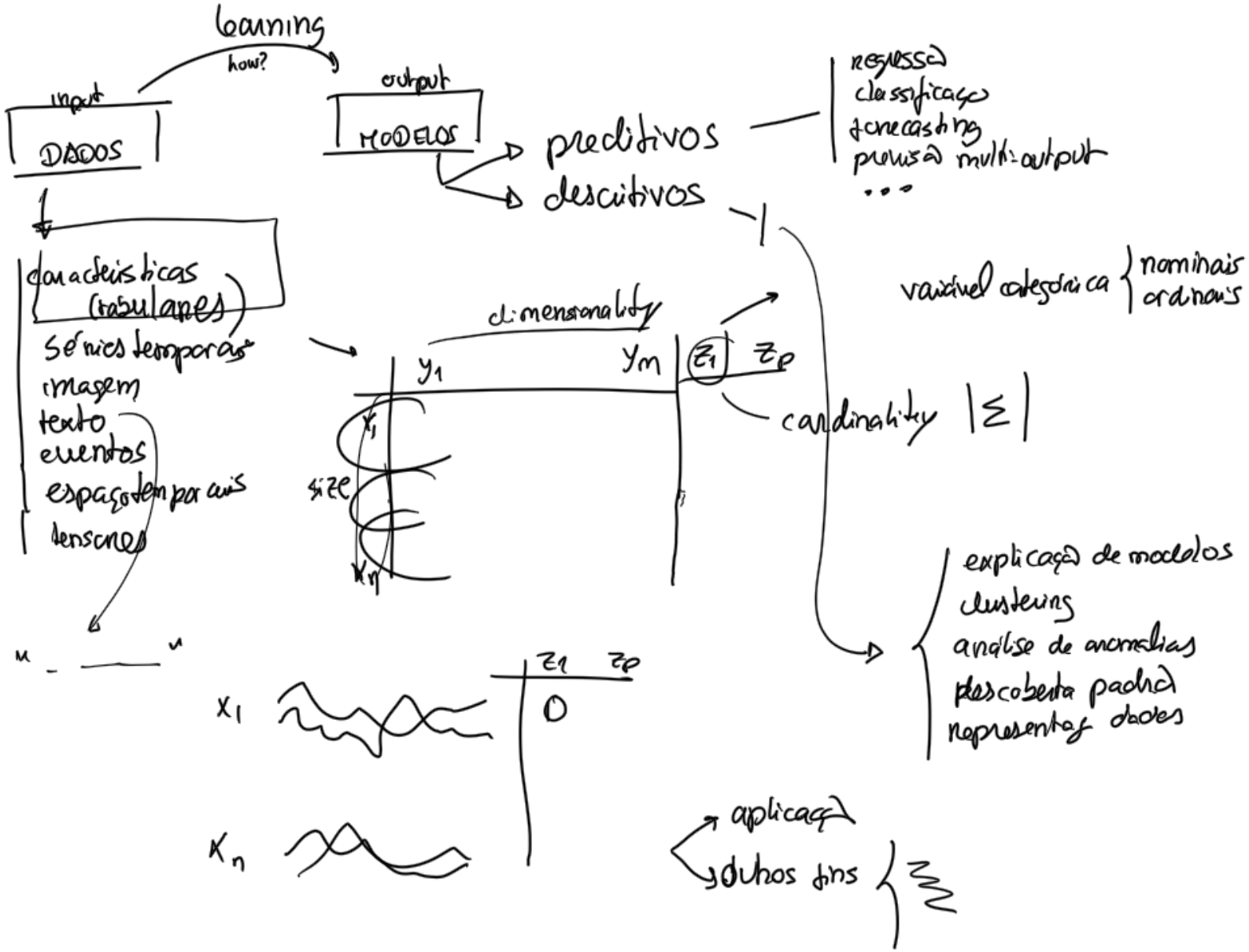
inteligência { adaptabilidade ← raza emoçã social

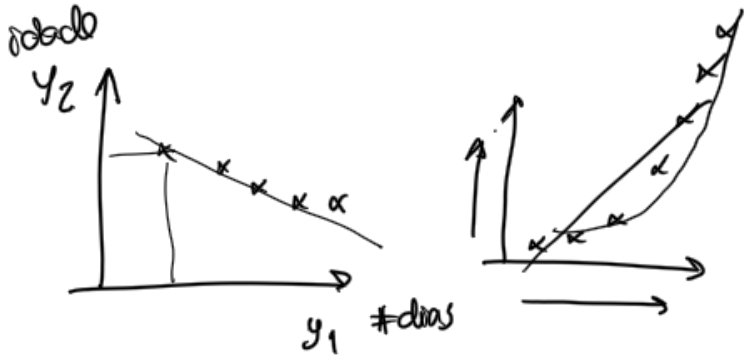
① IA
Aprend.
CD.

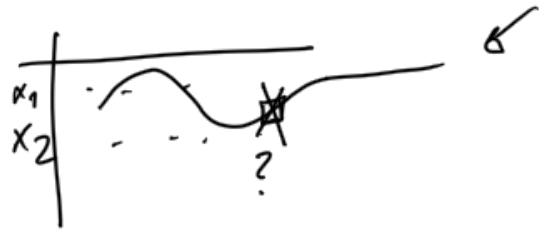


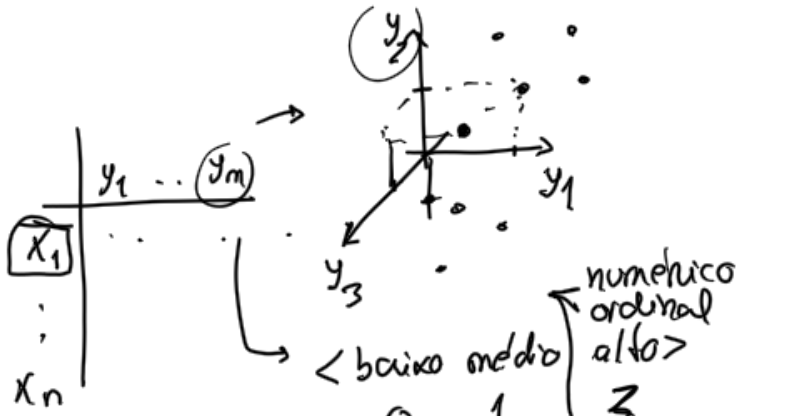
aplicaçã
↳ visã sistémica




learning
how?
input
DADOS
output
MODELOS
preditivos
descritivos
regressão
classificação
forecasting
previsão multi-output
...
características (tabulares)
séries temporais
imagem
texto
eventos
espaço-temporais
densidade
variável categórica
nominais
ordinais
dimensionality
y_1
y_m
z_1
z_p
cardinality $|\Sigma|$
size
x_1
x_n
explicação de modelos
clustering
análise de anomalias
descoberta padrão
representação dados
aplicação
dados dns

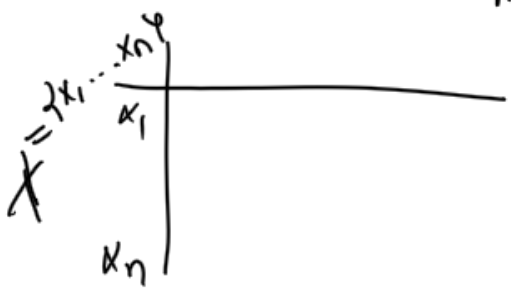




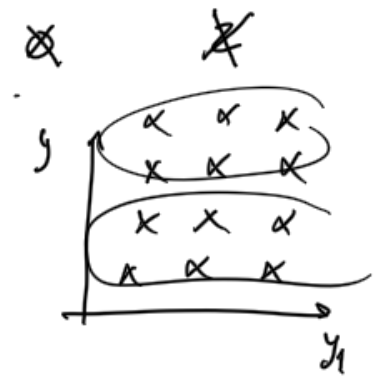


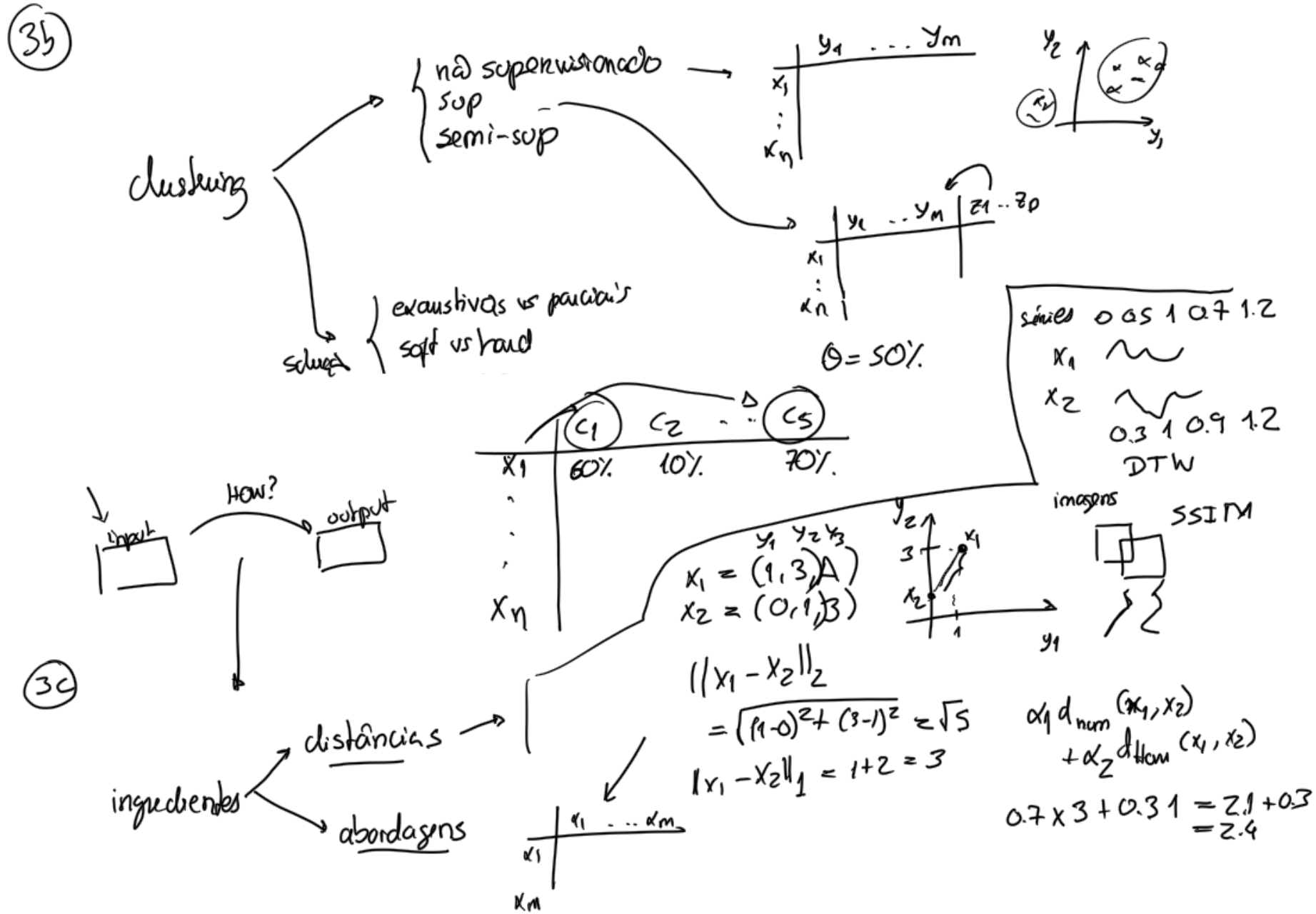
③ Clusterings

$C_K \subseteq X$

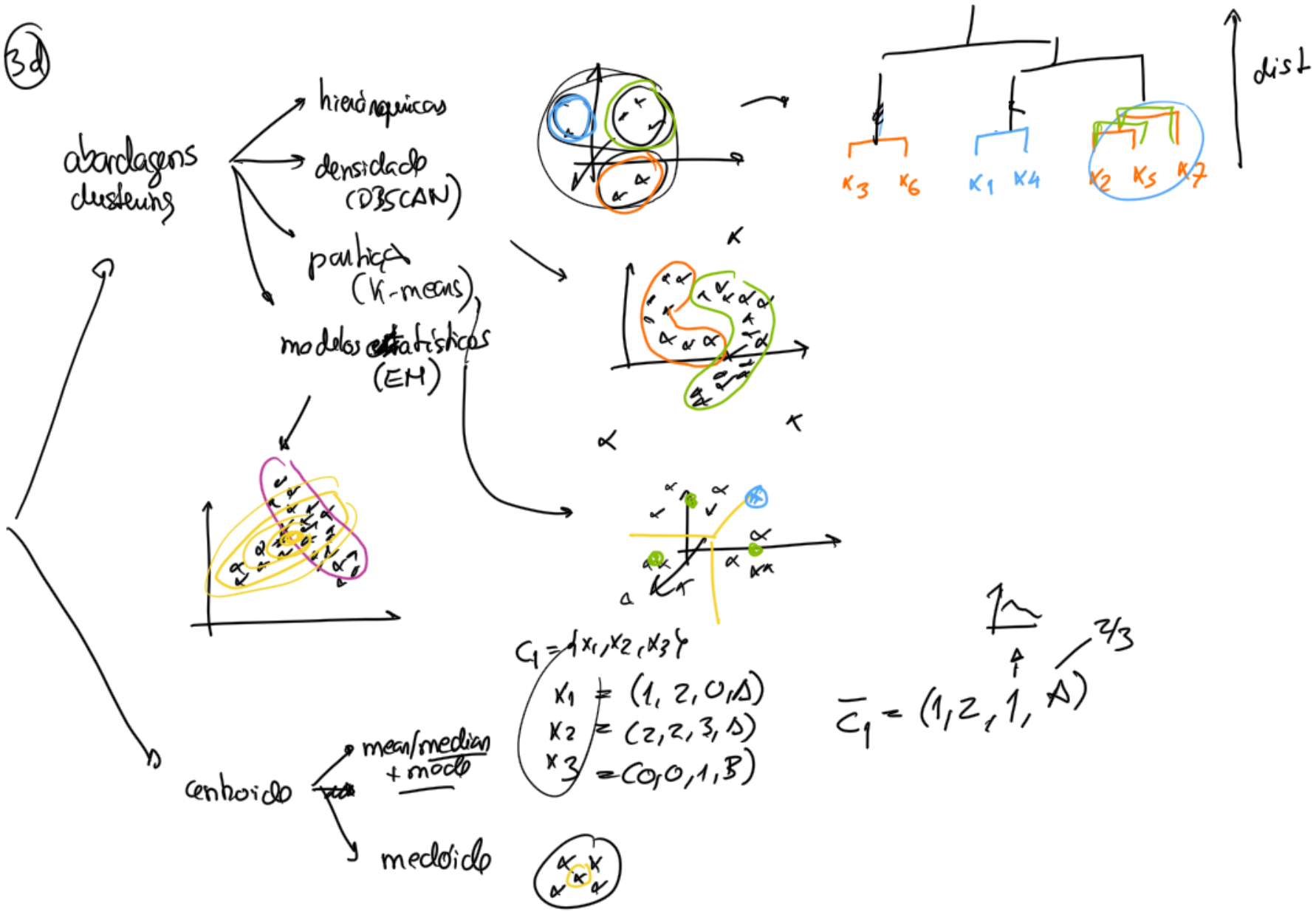


↗ coesão
↗ separação

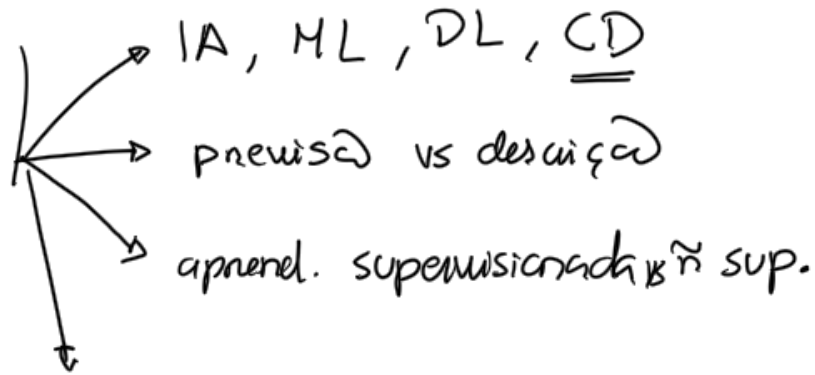


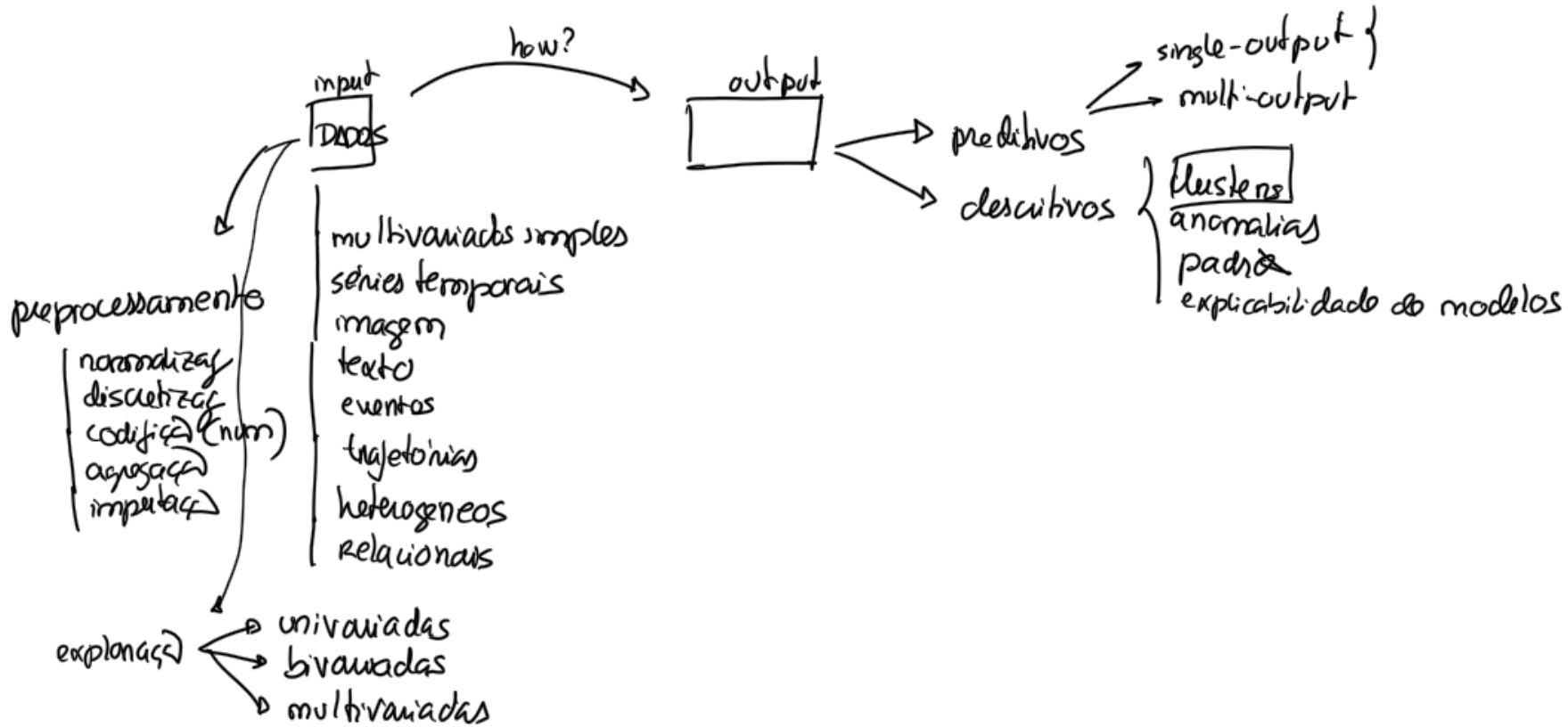

3b
clustering
não supervisionado
sup
semi-sup
exaustivas vs parciais
soft vs hard
θ = 50%
séries
DTW
imagens
SSIM
3c
How?
input
output
distâncias
ingredientes
abordagens

3d)


abordagens clustering
hierárquicas
densidade (DBSCAN)
particional (K-means)
modelos estatísticos (EM)
dist
x3 x6 x1 x4 x2 x5 x7
centroide
mean/median + mode
medoide
C1 = {x1, x2, x3}
x1 = (1, 2, 0, A)
x2 = (2, 2, 3, A)
x3 = (0, 0, 1, B)
C̄1 = (1, 2, 1, A)
2/3

① revisão


IA, ML, DL, CD
previsão vs descrição
aprend. supervisionada vs não sup.

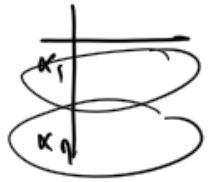

input
Dados
how?
output
preditivos
single-output
multi-output
descritivos
clusters
anomalias
padrão
explicabilidade de modelos
multivariados simples
séries temporais
imagem
texto
eventos
trajetórias
heterogeneos
relacionais
preprocessamento
normalizar
discretizar
codificar (num)
agregar
imputar
explicação
univariadas
bivariadas
multivariadas

(2)

clustering

- aplicação
  - marketing
  - bioinformática
  - educação
  - produtos/serviços
  - documentos/web pages
  - } "agrupar"

- fins
  - aquisição conhecimento
  - exploração → estrutura dados
  - apoiar aprend. / visualização / descrição
  - amostragem
  - compressão / denoising
  - eficiência



$\{C_1 .. C_K\}$

$C_j \subseteq X$

$\cup \{x_1 \dots x_n\}$

separação / coesão

kNN

	y_1	.	y_m	classe
x_1				A
				A
				B
x_n				B
				⋮

(3) ingredientes

- distâncias
  - mistas
- centróide
- clusterizações

$x_1 =$ (8, 5, A, C)

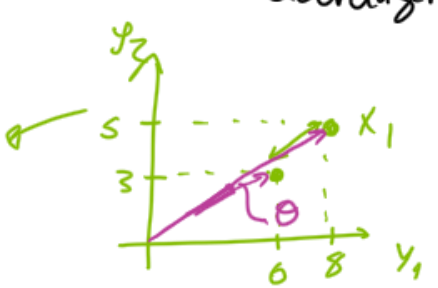
$x_2 =$ (6, 3, B, C)

$d(x_1, x_2) = \alpha_1 d_{Euc}(x_1, x_2) + \alpha_2 d_{Ham}(x_1, x_2)$

$\alpha_1 = 0.5$, $\alpha_2 = 0.5$

$d(x_1, x_2) = 0.5\sqrt{2^2+2^2} + 0.5(1+0) = 1.5 + 0.5$

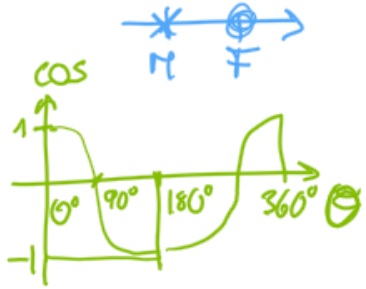
$\cos(x_1, x_2) = \frac{x_1 \cdot x_2}{\|x_1\| \|x_2\|}$

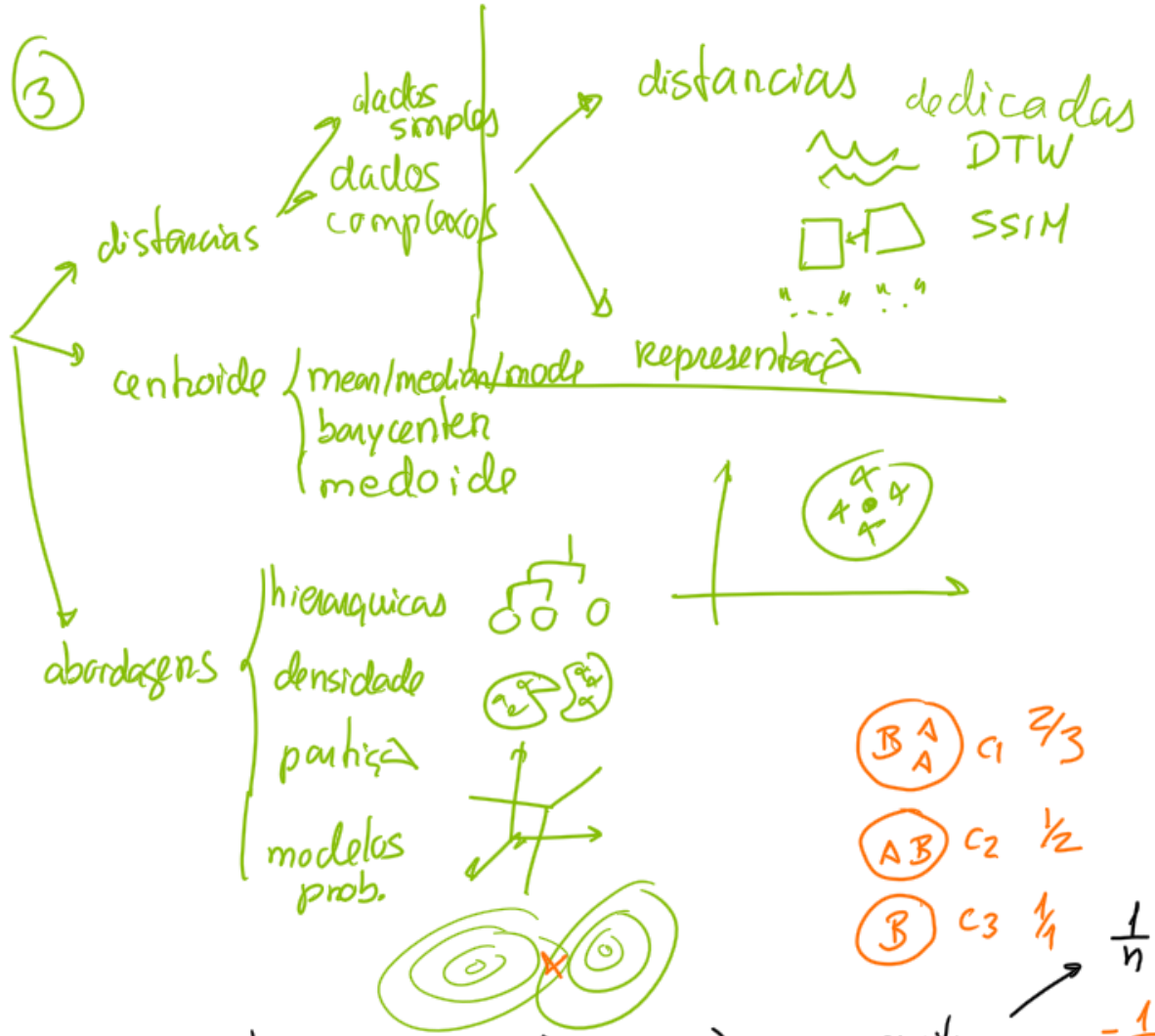


$= \frac{8\times6 + 5\times3}{\sqrt{8^2+5^2}\sqrt{6^2+3^2}} \approx 0.99$

(4) avaliação

- externa
- interna
  - separação
  - coesão
  - silhouette





4

avaliação
- externa
- interna

	G (true)	C (clusters)
x_1	A	C_1
x_2	A	C_1
x_3	B	C_1
x_4	B	C_2
x_5	B	C_3
x_6	A	C_2

- purity
- rand index

$$\frac{1}{n}\left(\sum_{C_k \in C} \max_j |C_k \cap G_j|\right)$$

$$= \frac{1}{6}\left(\max(2,1) + \max(1,1) + \max(0,1)\right)$$

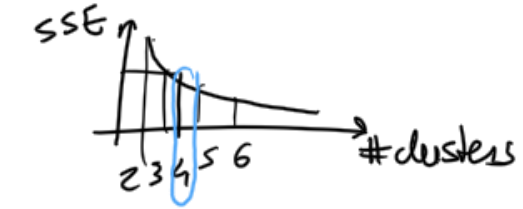
$$= \frac{1}{6}(2+1+1) = 2/3$$

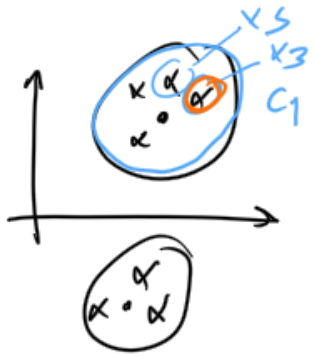
$$\frac{\#\text{concordant pairs}}{\#\text{pairs}} = \frac{1+6}{C_2} = \frac{7}{15}$$

(4)

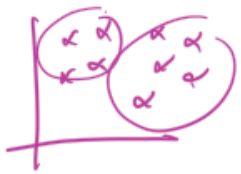
avaliação → interna → coesão, separação, silhouette

$$SSE = \sum_{C_k} \sum_{x_i \in C_k} d(x_i, \bar{c}_k)^2$$

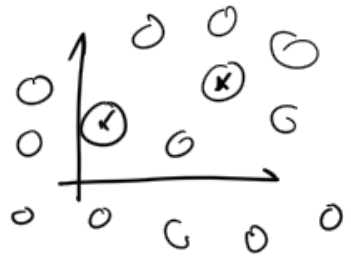




(1, 2, Δ)
(2, 4, Δ) ⊢ (1.5, 3, Δ)



n = 100 observações
K = 100 clusters



obs
cluster
solução

$$s(x_i) = 1 - \frac{a(x_i)}{b(x_i)}$$

$\frac{b}{a} - 1$ (nº obs) $a < b$

	x_1	x_2	x_3	x_4	
x_1	0	2	1	3	
x_2	0	3		2	C_1
x_3	0			4	
x_4				0	C_2

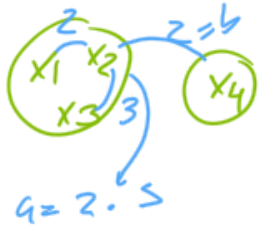
$s(x_1) = 1 - \frac{1.5}{3} = 0.5$

$s(x_2) = \frac{2}{2.5} - 1 = -0.2$

$s(x_3) = 1 - \frac{2}{4} = 0.5$

$s(x_4) = 1 - 0 = 1$

$s(C_1) = \frac{s(x_1) + s(x_2) + s(x_3)}{3} = 0.2(6)$

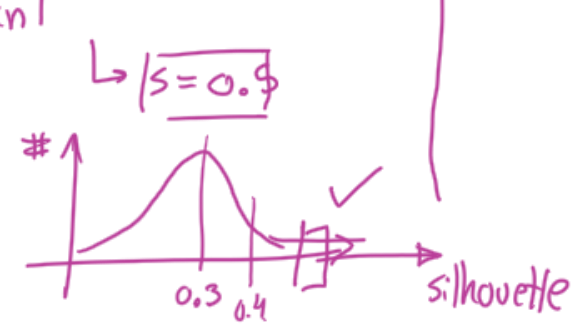


sig. estatística — p-value 3%

$y_1 \dots y_4$ → 1000

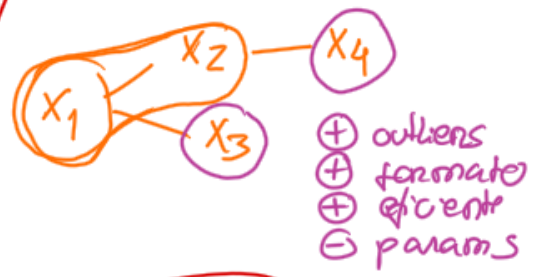
x_1 … x_n

↳ $s = 0.5$



(5) aglomerações

- hierárquico (aglomerativo)
- densidade (DBSCAN) → $K = 2$, $\varepsilon = 2.5$
- K-means
- modelos prob. (EM)



⊕ outliers
⊕ formato
⊕ eficiente
⊖ params

	x_1	x_2	x_3	x_4
x_1	0	2	1	3
x_2		0	3	2
x_3			0	4
x_4				0

K-Means

	y_1	y_2
x_1	0	0
x_2	1	0
x_3	0	2
x_4	2	2

$\bar{c}_1 = (2,0)$
$\bar{c}_2 = (2,1)$

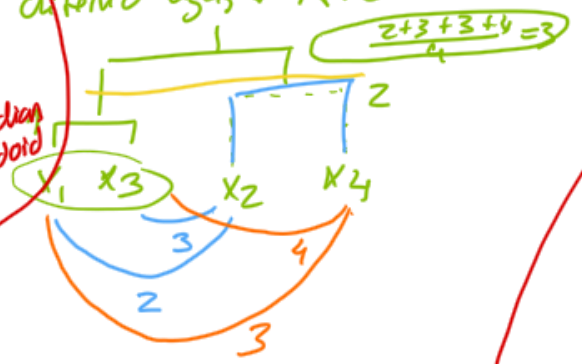
↳ 1ª it. $C_1 = \{x_1, x_2\}$, $C_2 = \{x_3, x_4\}$
$\bar{c}_1 = (0.5, 0)$
$\bar{c}_2 = (1, 2)$ → 2ª it

⊕ eficiente
⊖ globulares
⊖ K fixo
⊖ outliers → K-medians, K-medoid

a) critério ligação MIN

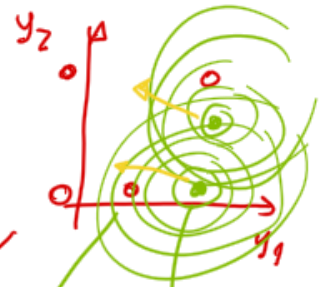
⊕ formato
⊖ tamanhos

critério ligação AVG



critério ligação MAX

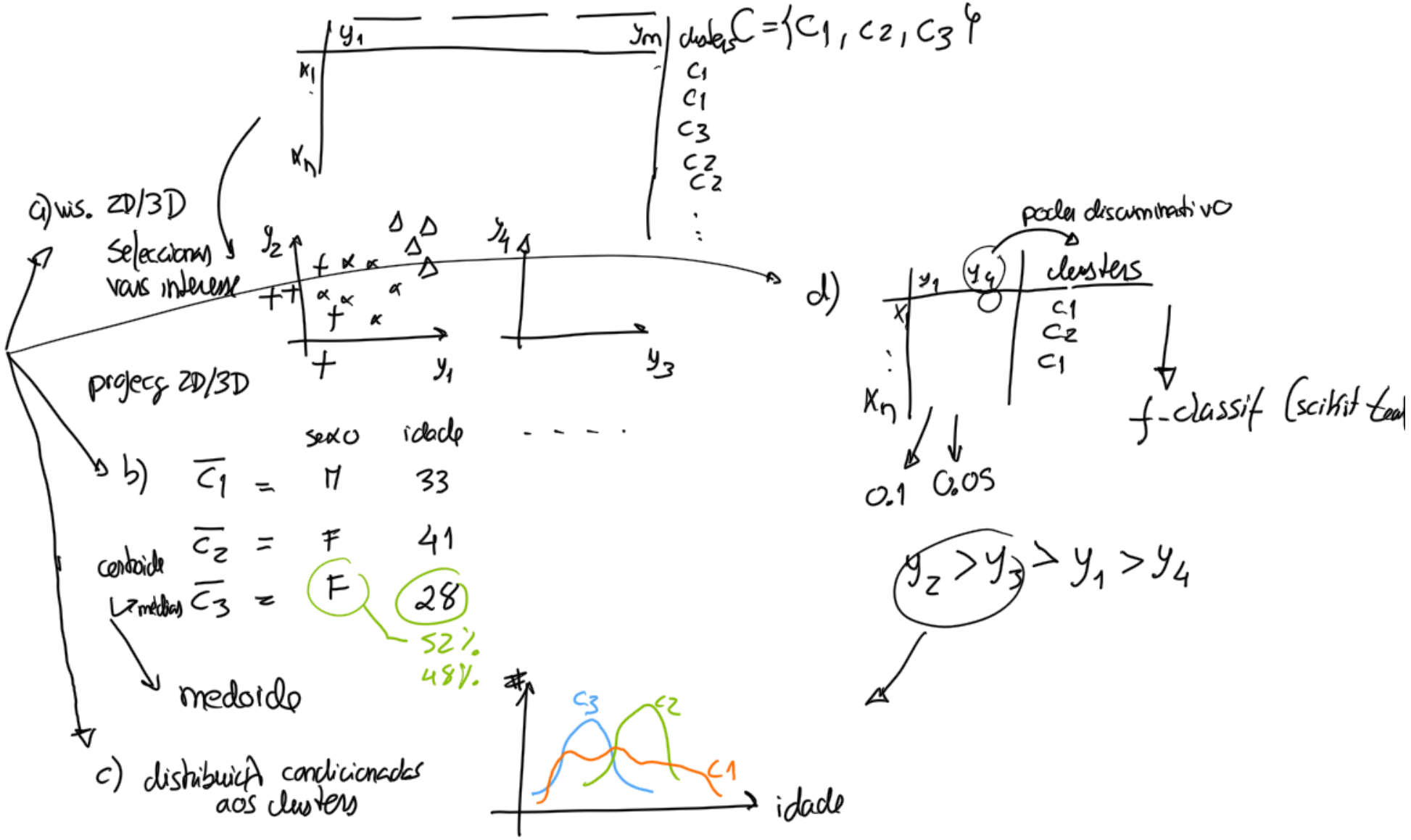
⊖ globulares
⊕ unif. tamanhos

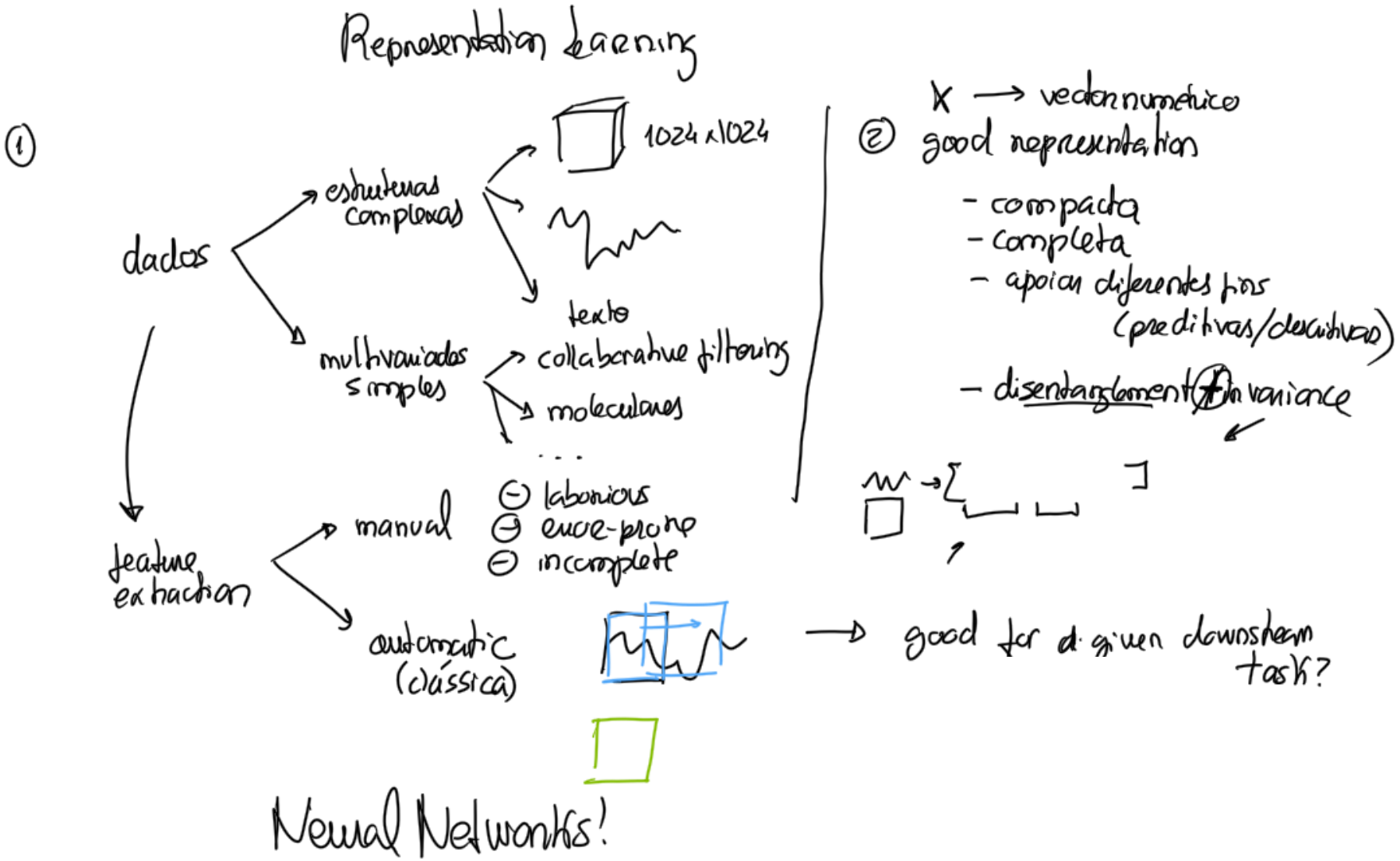


$p(x_2 | c_1)$

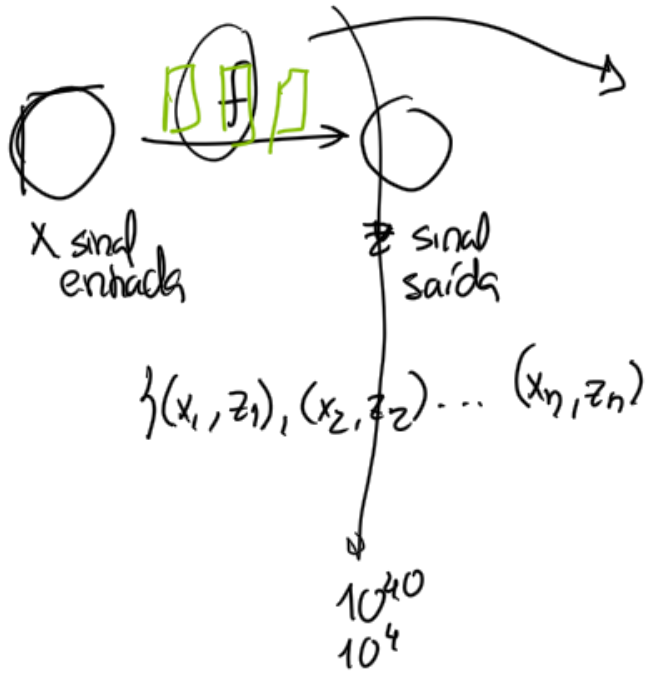
⊕ soft/prob
⊖ pressupostos

(6) visualização clustering




Representation learning
①
dados
estruturas complexas
1024 x 1024
texto
multivariados simples
collaborative filtering
moleculares
...
feature extraction
manual
⊖ laborious
⊖ error-prone
⊖ incomplete
automatic (clássica)
Neural Networks!
② good representation
- compacta
- completa
- apoia diferentes tipos (preditivas/descritivas)
- disentanglement ≠ invariance
→ good for a given downstream task?

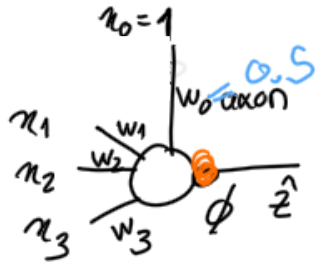
③



neurónio (rd)

$\vec{x} = (0, 2, 1)$ (x_1 x_2 x_3)

$z = 1$



$\hat{z} = \phi\left(\sum_{i=0}^{n} w_i x_i\right)$

$w = (1, -1, 2)$ (w_1 w_2 w_3)

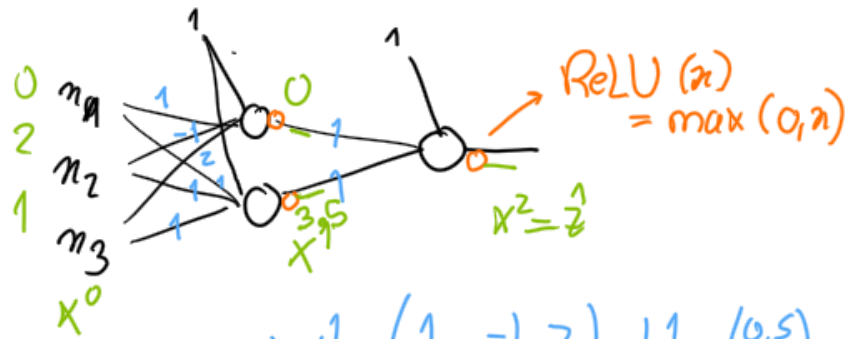
$\hat{z} = w_0 x_0 + w_1 x_1 + w_2 x_2 + w_3 x_3$

$= 0.5 \times 1 + 1 \times 0 - 1 \times 2 + 2 \times 1$

$= 0.5$

$|z - \hat{z}| = |1 - 0.5| = 0.5$

Redes neuronais



$W^1 = \begin{pmatrix} 1 & -1 & 2 \\ 1 & 1 & 1 \end{pmatrix}$ $b^1 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$

$W^2 = (1\ 1)$ $b^2 = (0.5)$

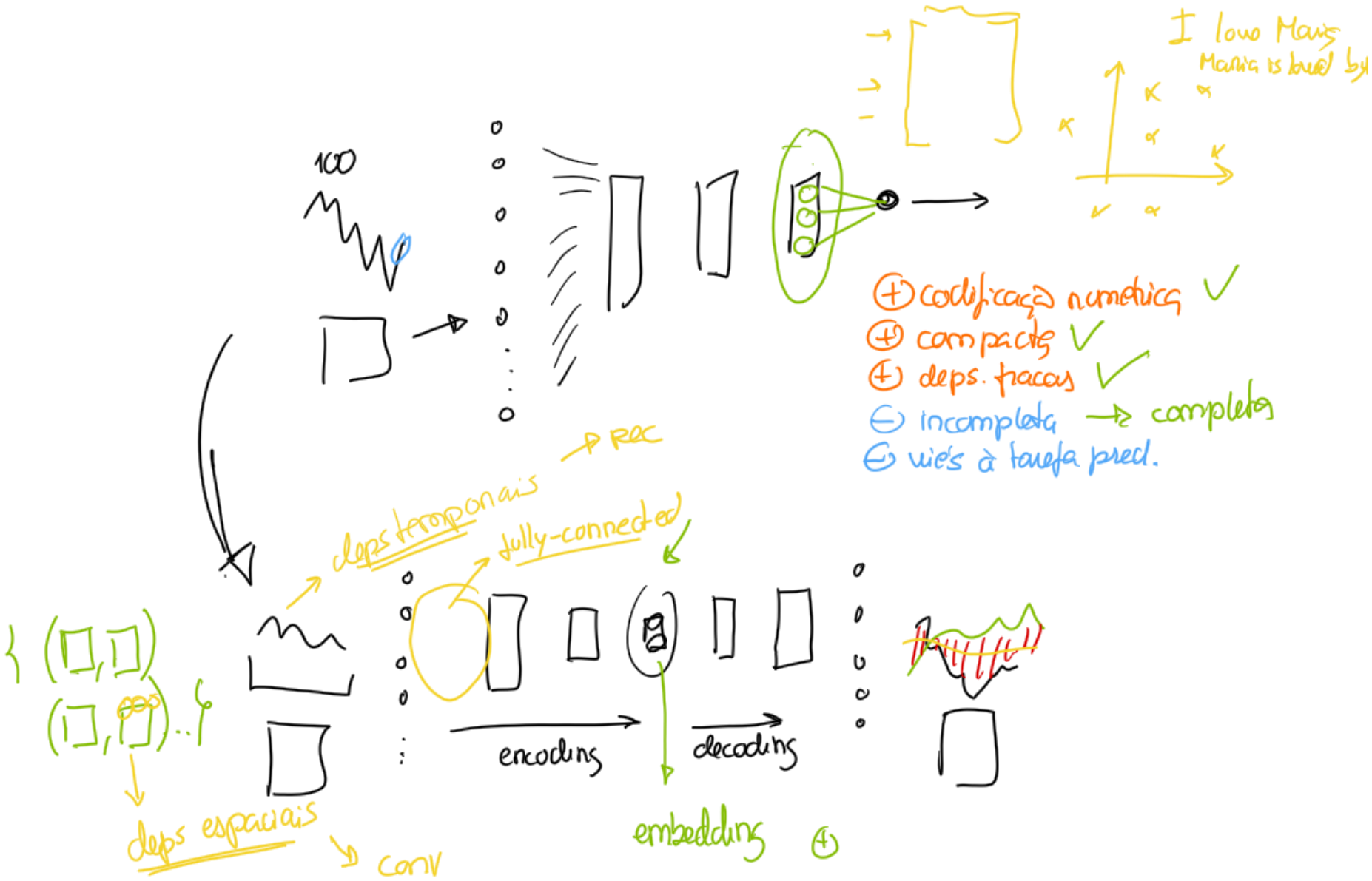
$x = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$

$x^1 = \phi(W^1 x^0 + b^1)$

$= \phi\left(\begin{pmatrix} 1 & -1 & 2 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} + \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}\right)$

$= \phi\left(\begin{pmatrix} 1 \\ 3 \end{pmatrix} + \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}\right) = \phi\begin{pmatrix} -0.5 \\ 3.5 \end{pmatrix} = \begin{pmatrix} 0 \\ 3.5 \end{pmatrix}$

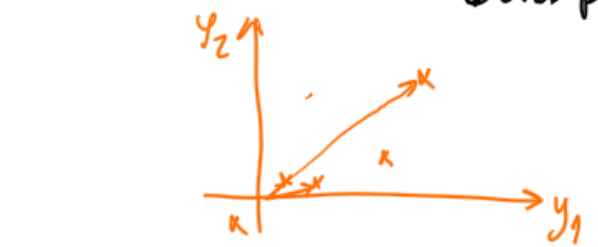
$\hat{z} = x^2 = \phi(W^2 x^1 + b^2) = \phi\left((1\ 1)\begin{pmatrix} 0 \\ 3.5 \end{pmatrix} + (0.5)\right) = \phi(4) = 1$


I love Maria
Maria is loved by
100
⊕ codificação numérica ✓
⊕ compacta ✓
⊕ deps. fracas ✓
⊖ incompleta → completa
⊖ viés à tarefa pred.
deps temporais
rec
fully-connected
deps espaciais
conv
encoding
decoding
embedding

0 Recap
- introdução
  - aprend sup vs n sup
  - descrição vs previsão
- estatística descritiva
- clustering
  - abordagens + distâncias
  - avaliação { externo / interno
- representação de dados

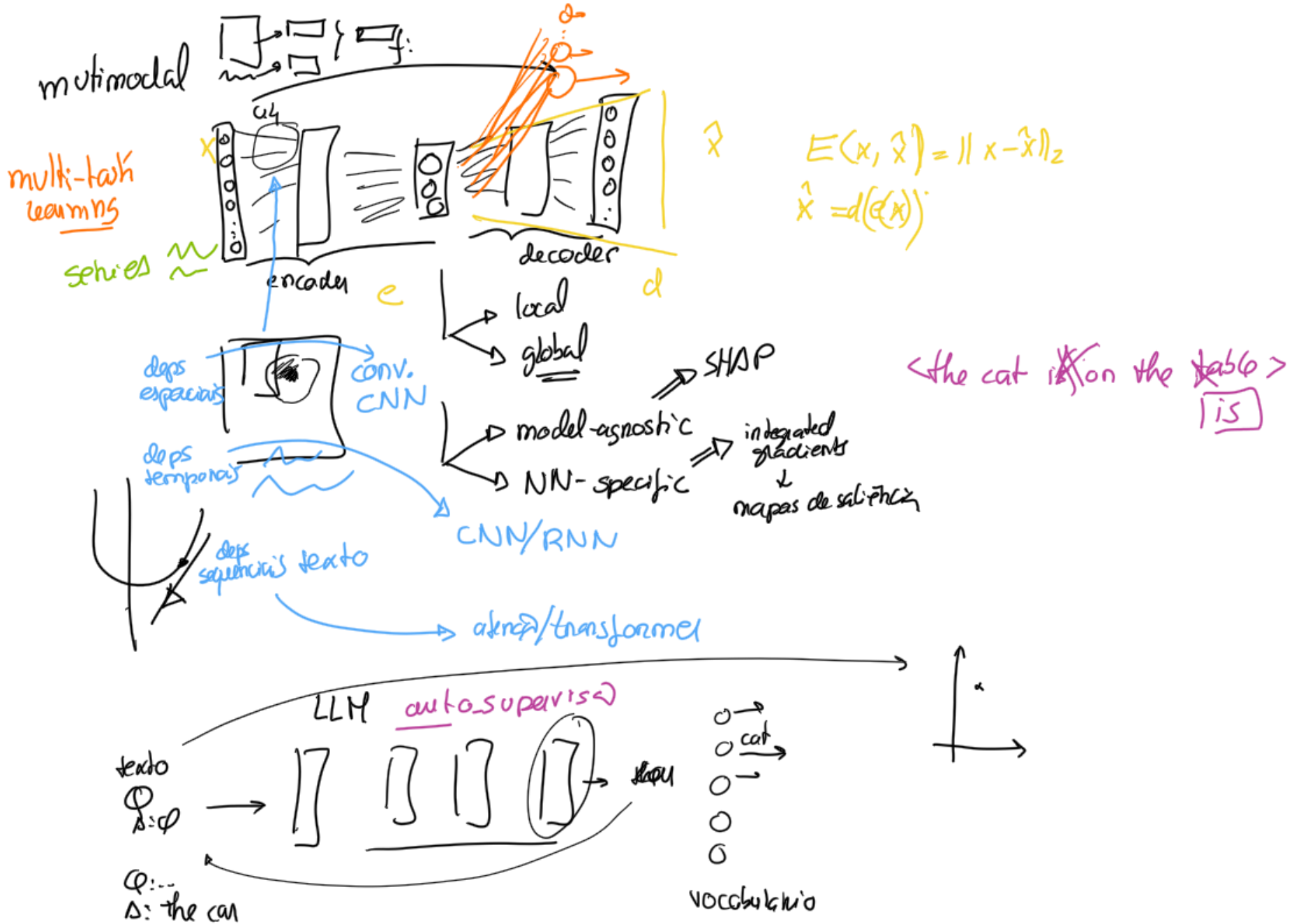
1
- estruturas complexas (extração características)
  - manual ∝
  - domain-driven automated extraction
  - advanced (domain-agnostic) automated extraction
- NNs!

representação de dados
- sucintas/compactas
- completa/informativa
- apoiam as tarefas aprend.
- outras propriedades { invariância / desintrelacionamento / ⊥

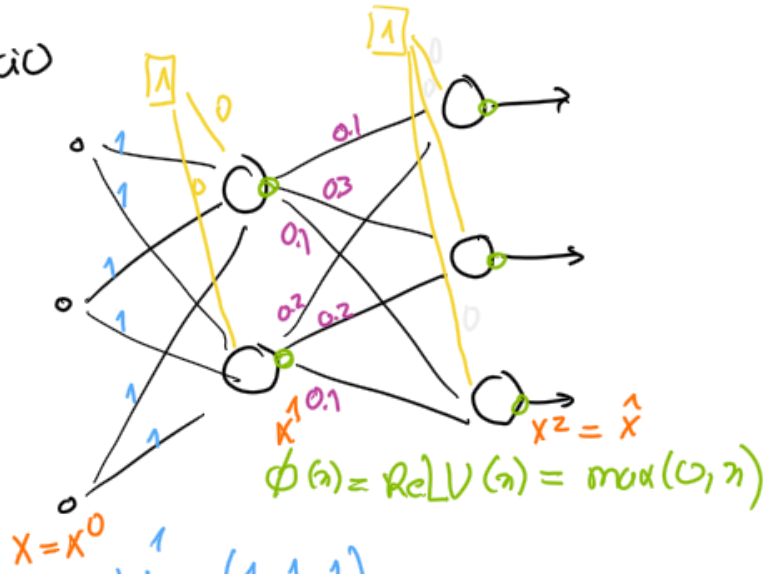


cosine

<0 0 1 0 0.3
<0 0 2 0 0.6


multimodal
multi-task learning
series
encoder
decoder
$E(x, \hat{x}) = \|x - \hat{x}\|_2$
$\hat{x} = d(e(x))$
local
global
SHAP
model-agnostic
integrated gradients
NN-specific
mapas de saliência
deps espaciais
conv. CNN
deps temporais
CNN/RNN
deps sequenciais texto
atenção/transformer
<the cat is on the table>
is
LLM
auto-supervisão
texto
cat
vocabulário

④ exercício



$$W^1 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

$$b^1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$W^2 = \begin{pmatrix} 0.1 & 0.2 \\ 0.3 & 0.2 \\ 0.1 & 0.1 \end{pmatrix}$$

$$b^2 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Representação da observação $x = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$

$$x^1 = \phi(W^1 x^0 + b^1)$$

$$= \phi\left(\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right) = \phi \begin{pmatrix} 4 \\ 4 \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \end{pmatrix}$$

ReLU

reconstrução

$$\hat{x} = x^2 = \phi(W^2 x^1 + b^2)$$

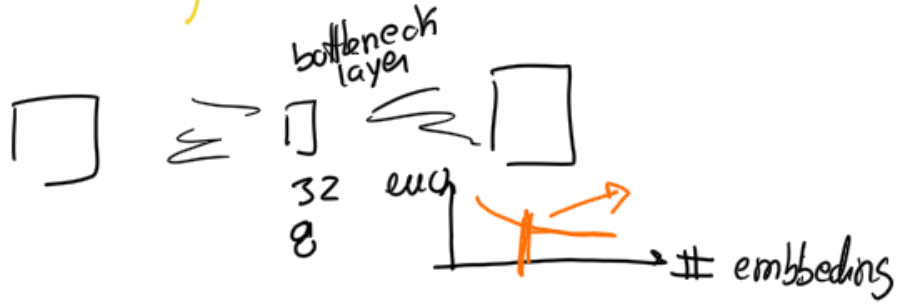
$$= \phi\left(\begin{pmatrix} 0.1 & 0.2 \\ 0.3 & 0.2 \\ 0.2 & 0.1 \end{pmatrix} \begin{pmatrix} 4 \\ 4 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right) = \phi \begin{pmatrix} 1.2 \\ 2 \\ 0.8 \end{pmatrix} = \begin{pmatrix} 1.2 \\ 2 \\ 0.8 \end{pmatrix}$$

erro $\|x - \hat{x}\|$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2} = \sqrt{\frac{1}{3}(0.2^2 + 0^2 + 0.2^2)} = 0.16$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |x_i - \hat{x}_i|$$

$$= \frac{1}{3}(0.2 + 0 + 0.2) = 0.13$$



⑤ transformações

before:  → <...>
complexos rep. numéricas

now <...> → ϕ <...>

univariadas

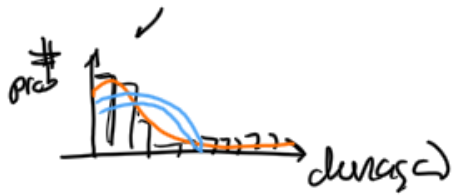
	tempo hospitalização	log
x_1	0.1	
	1	
	⋮	
x_n	22	

x_1	y	$1+\log_2(y)$
	1	— ①
	8	— $1+\log_2 8$
		$1+3=4$
	32	— $1+\log_2 32$
		$1+5=6$

multivariadas

$\mathbb{R}^3 \xrightarrow{\phi} \mathbb{R}^2$

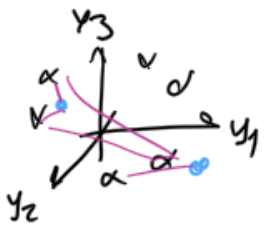
$(y_1, y_2, y_3) \mapsto (y_1 y_2, y_1 + y_3)$



log duração

$\not\mapsto (y_1, y_2, y_3, y_1 + \alpha y_3) \in \mathbb{R}^4$

Kernel



ϕ

	y_1 ... y_m
x_1	
⋮	
x_n	

$\mapsto$

	c_1 ... c_p
x_1	
⋮	
x_n	

$p \ll m$

7)

redução dimensionalidade

- feature selection $y' \subseteq Y$
  - sup — poder preditivo (ganho informação)
  - unsup — variabilidade
- aprendizagem
  - árvores decisão — podagem
  - Redes neuronais — regularização
  - ...
- subspace/subsample selection
  - $\exists$
    - linear
      - linear → LDA
      - não linear → PCA
    - não linear

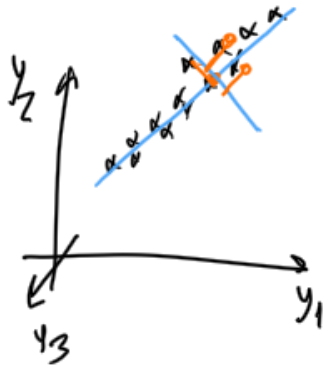
$y_1 \dots y_m \quad c_1 \dots c_p$

$\| \quad f_1(y_1, \dots y_m) = \alpha_1 y_1 + \alpha_2 y_2 + \dots \alpha_m y_m$

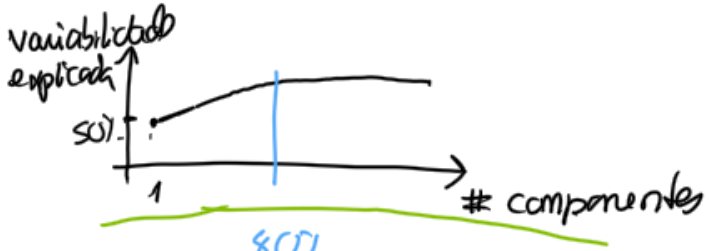
(8)

PCA

MMAP
tSNE



$y_1 \dots y_{1000} \rightarrow c_1 \qquad c_{1000}$



80%
90%
95%
98%

	y_1	y_2	y_3
x_1	1	1	1
x_n			

$$\begin{cases} c_1 = 2y_1 + y_2 - 2y_3 \\ c_2 = y_1 - 2y_2 \end{cases}$$

$\rightarrow v_1 = \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}$

$\rightarrow v_2 = \begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}$

	c_1	c_2
x_1	1	-1

$2+1=2$ $\qquad 1-2+0$

$$C = \begin{matrix} y_1 \\ y_2 \\ y_3 \end{matrix} \begin{pmatrix} & y_1 & y_2 & y_3 \end{pmatrix}$$

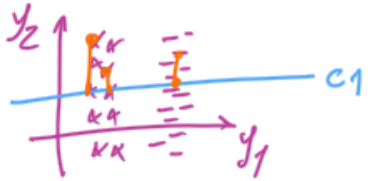
v_1	$\lambda_1 = 7$	$\frac{\lambda_1}{\sum_i \lambda_i} = \frac{7}{7+2+1} = \frac{4}{7} \approx 70\%$
v_2	$\lambda_2 = 2$	
v_3	$\lambda_3 = 1$	

$$U = \begin{pmatrix} 2 & 1 \\ 1 & -2 \\ -2 & 0 \end{pmatrix}$$

$$x' = U^T x = \begin{pmatrix} 2 & 1 & -2 \\ 1 & -2 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$x = (U^T)^{-1} x' = U x' = \begin{pmatrix} 2 & 1 \\ 1 & -2 \\ -2 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix}$$

(9) LDA



(16)

Análise Anomalias

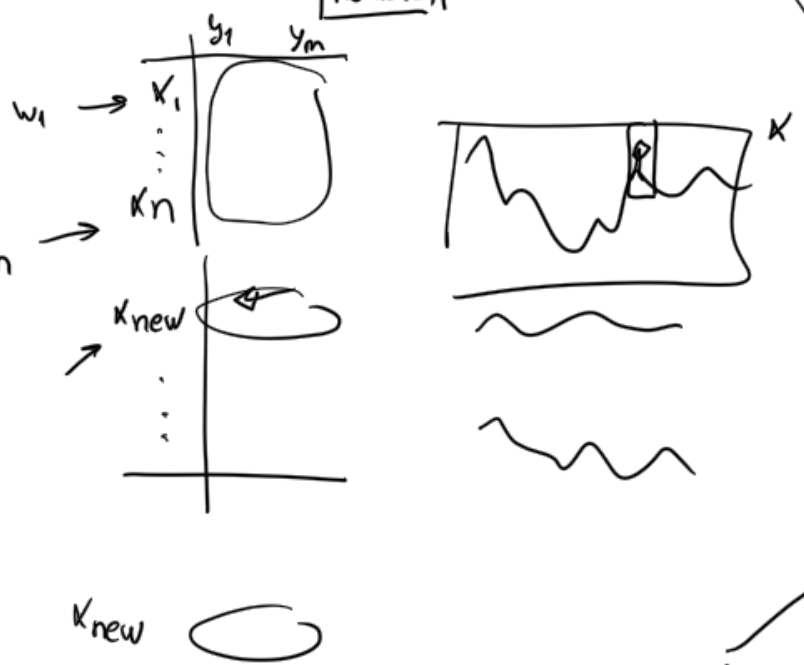
anomalia
outlier
exceção
desvio
≠
novidade

- univariado
- multivariado

$y_1 \dots y_m$

68%

$M \pm 2\sigma \rightarrow 95\%$
$3\sigma \rightarrow 99.7\%$

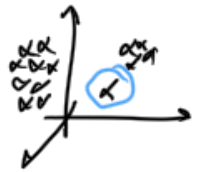


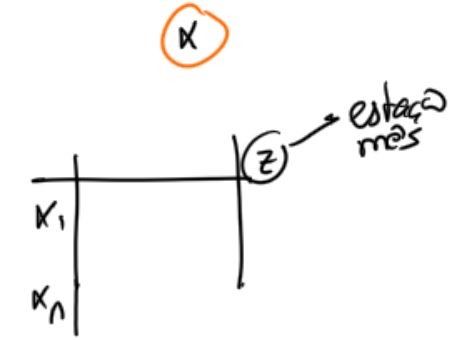
- deteção
- tratamento
  - remover
  - manter
  - corrigir
  - pesar

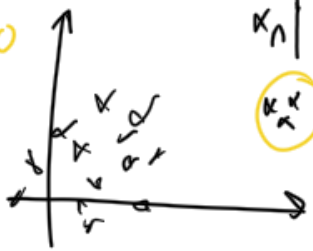
- fins
  - aquisição conhecimento
  - explicação
  - apoiar previsão/deteção (preprocessamento)
  - ...
- aplicação
  - remoção ruído
  - fraude
  - cibersegurança
  - saúde
  - avarias
  - sensores
  - consumos
    - produtos/recomend
    - aprovar crédito
    - ...

11

anomalias tipos
- globais
- locais
- normal
- contextual
- coletiva



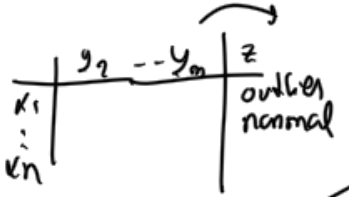




- ⊖ separar outliers ↔ ruído
- ⊖ interpretabilidade
- ...

12

paradigmas
- sup
- semi-sup
- ñ sup

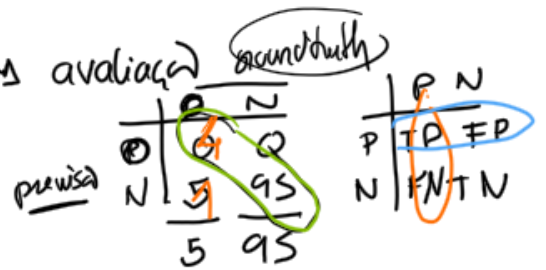


$f: (y_1 \dots y_m) \rightarrow z$ classificação regressão

x_{new}

ñ balanceamento
- sampling
- learning
- ...

avaliação



R = 80%.

$R = 0.8 \mid P = 0.2 \mapsto 0.5$

$F_1 \rightarrow F_\beta = z$

F-measure $_\beta$

$prec = \frac{TP}{TP + FP}$

$recall = \frac{TP}{TP + FN}$

$acc = \frac{TP + TN}{all}$

13

semi-sup

normais + outliers

outliers (unsup)

normais (unsup)

pseudo-anotar

+ pesagem obs

14

unsup

modelos prob

~~distâncias/densidade~~

clustering

deep learning (NN)

univ

multiv

(-) pressuposto

K vizinhos

abaixo de um ε

(DBSCAN)

[dist média aos k vizinhos]

$p(\underline{x}) = p(x \mid y_1, y_2)$

$= p(x \mid y_1) \times p(x \mid y_2)$

$= 0.2 \times 0.8 = 0.16$

(1,3)

Descoberta de padrões

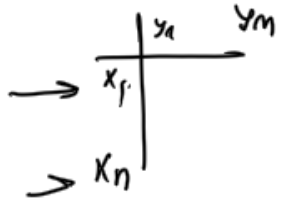
31/1/2026

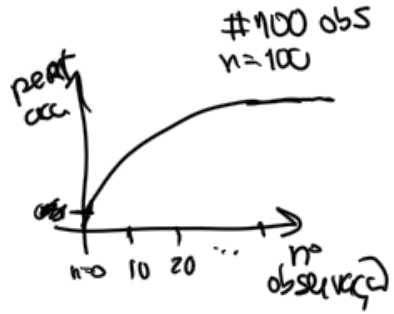
0

até aqui
- introdução (IA → ML → descrição)
- estatística descritiva | univariada / bivariada
- clustering { distâncias, abordagens, avaliação, visualização → ag. conhecimento
- anomalias { tipos (globais, locais, contextuais, coletivas); abordagens → sup, semi-sup, ñ sup; tratamento | remover, manter, pesar, corrigir
- representação de dados (complexos) → domain-driven; domain-agnostic → embeddings (redes) – codificação numérica
- transformação → univariada { log, norm, seleção de vars; multivariada → lineares { PCA, LDA; ñ lineares (redes), MMAP
- avaliação → aprender com/sem transformação { pred, desc; propriedades da representação | tamanho, % erro de reconstrução; ...

riscos (generalização)
- overfitting
- underfitting

| pred | drivenness dos redes
↓
desc

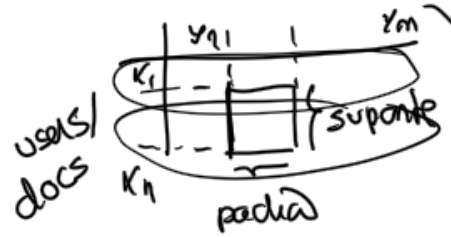




① descoberta de padrões

prever → decisão
descrever → aq. conhecimento

clusters → regularidades globais → regularidades locais → padrões



produtos/contextos

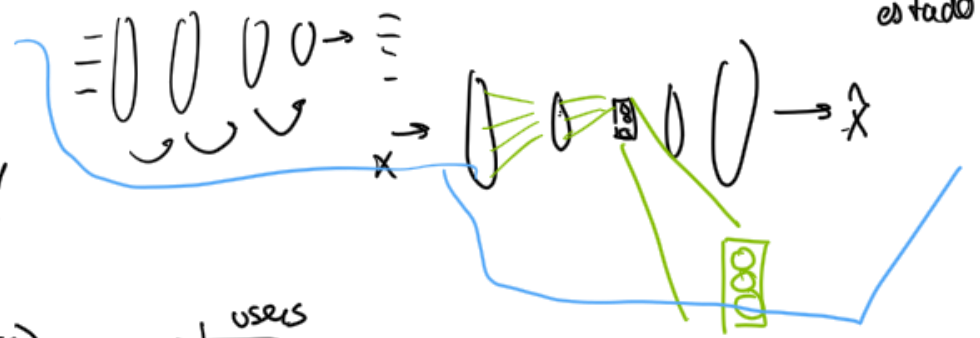
mutação ∧ ∧ sexo ≠
∧ bmanc < θ ⇒ AD

estado B ∧ $y_{saúde}^{C}$ ⇒ $y_{saúde}^{D}$

②

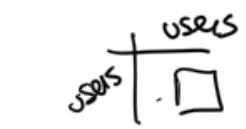
pattern recognition

padrão → latente
padrão → explícito
pattern discovery mining

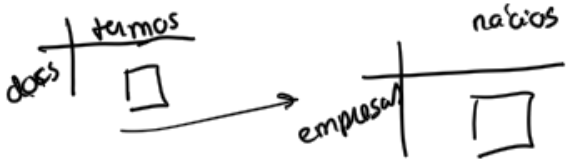


③

aplicação → recomendação
redes sociais
mercados
web
pesquisa
economia/finanças



fraldas ∧ cerveja

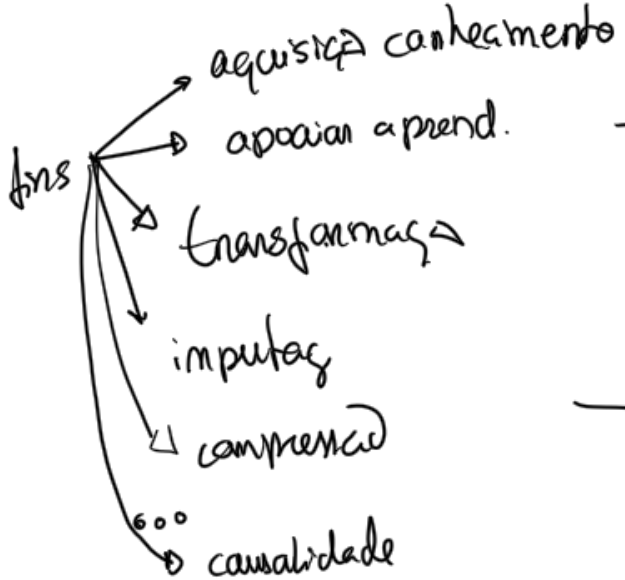


biotec
saúde
nanociência

↑A ∧ ↓B ⇒ ↗ composto C

~~lost~~ A ∧
slico B ⇒

④



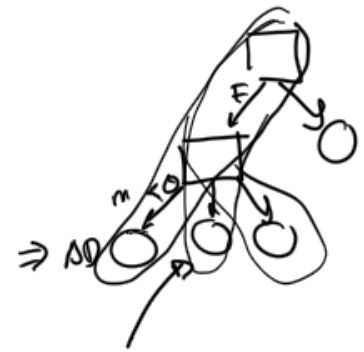
	p_1 ... p_m	t
x_1	1 0 0 10	
x_n		

prev
dec

	y_2 y_4 y_{12}
x_3	A B D
x_5	? B D
x_6	A B D

$B \Rightarrow C$

⑤



cj padrões
ordenados / agrupados / seriados
(hierárquicos)

transaccionais
multiv. simples
⋮

itemsets $x_1 = \{A, C, F\}$
$x_2 = \{B, C\}$

	A	B	C	D	E	F
x_1	1	0	1	0	0	1
x_2	0	1	1	0	0	0

itemset seqs $x_1 = \langle \{A, C, F\}, \{B, F\} \ldots \rangle$

(6)

padrão
- informativo → simples
- preditivo / descritivo → regras de associação (padrão discriminatórios)

$y_1 < \theta_1 \wedge y_2 > \theta_2$

$(A \wedge B) \vee (A \wedge C)$

$A \cap D \Rightarrow E$

$y_A = 1 \cap y_D = 1 \Rightarrow y_E = 1$

Sistema → $y_1 \dots y_m$

Sistemas → x_1, x_7

frequência/suporte

$P(\sigma \cap \text{altura} > 1.7) = 0.6$

$P(\text{saldas} \cap \text{cerveja}) = 0.001$

inesperadamente frequentes

$P(A \cap B) \gg P_{null}(A \cap B)$

$N = 1000$

água = 300 $\quad P(A) = \frac{300}{1000} = 0.3$

batata = 100 $\quad P(B) = 0.1$

água ∩ batata = 50 (35) $\quad P(A \cap B) = 0.05$

$P_{null}(A \cap B) \overset{A \perp B}{=\!=} P(A) P(B) = 0.3 \times 0.1 = 0.03$

Significância estatística (π ocorre por chance)

$n \sim Bin(N = 1000, p_{null} = 0.03)$

$P(n \geq 50 \mid N = 1000, p_{null} = 0.03) = 2.4 \times 10^{-4}$ $\quad < 0.05$ tem sig. estatística

$P(n \geq 35 \mid \dots) = 0.15$

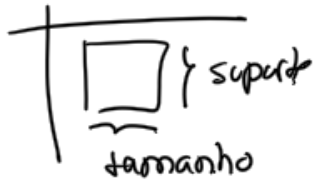
padrões simples → freq. itemsets (mixtend), biclusters, ...

métodos/procuras → exaustivas, aproximadas

p1, p2 — top-10 padrões

1) frequência/suporte 0.1
2) significância estatística 0.05
3) tamanho 3

p-value = $a E^{-10}$



⑦ regras de associação

R: A ⇒ B

antecedente — consequente

cerveja ⇒ fraldas

fraldas ⇒ cerveja

água P(A) = 0.3

batatas P(B) = 0.1

P(A∩B) = 0.05

4) confiança $R = \frac{sup(A \cap B)}{sup(A)} \geq \frac{P(A \cap B)}{P(A)} = \frac{0.05}{0.3} = 16.6\%$ (0.7)

sup(A∩B) 10

água → batatas → $\frac{P(A \cap B)}{P(A)} = \frac{10}{300} = 3.(3)\%$

batata → água → $\frac{P(A \cap B)}{P(B)} = \frac{10}{100} = 10\%$

5) Lift $(R) = \frac{conf(R)}{P(B)}$ (1.3, 1.4) sup(A∩B) = 10

→ $LiftR = \frac{\frac{P(batatas \cap água)}{P(batatas)}}{P(água)} = \frac{0.1}{0.3}$

= 1.3

R: Batatas ⇒ Água

$\frac{\frac{P(A \cap B)}{P(A)}}{P(B)} = \frac{P(A \cap B)}{P(A)P(B)} \overset{A \perp B}{=} \frac{P(A)P(B)}{P(A)P(B)} = 1$

< 1 A e B associação negativa

> 1 A e B tem associação forte

= 1 A e B idem associação

(8)

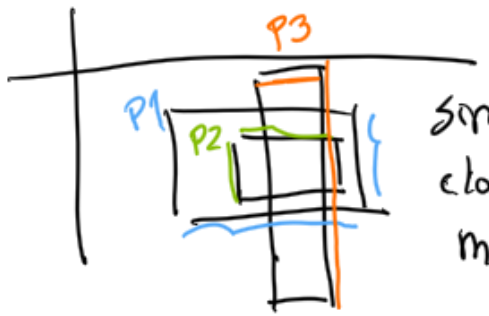
padrão relevante
- poder informativo / discriminativo
  - ↳ p-value
  - ↳ lift
- novidade / ≠ trivialidade
  - ↳ conta conhecimento existente
- accionabilidade → guiar decisão
- dissimilaridade

incorporar conhecimento
- ↳ restrição
  - inclusão/exclusão
  - alvo no consequente
    - ↳ lista supervisionada
  - sum (valor cabaz) ≥ 100

y_1
~~A~~
~~A~~
B
~~C~~
~~A~~
B

dummify

y_1A	y_1B	y_1C
0	0	0
0	1	0
0	0	0
0	0	1
0	0	0
1	0	0



simples: P1, P2, P3
closed / fechados: P1, P3
máximas: P1

diagnóstico

$y_1 \dots y_m \mid z_1 \dots z_m$

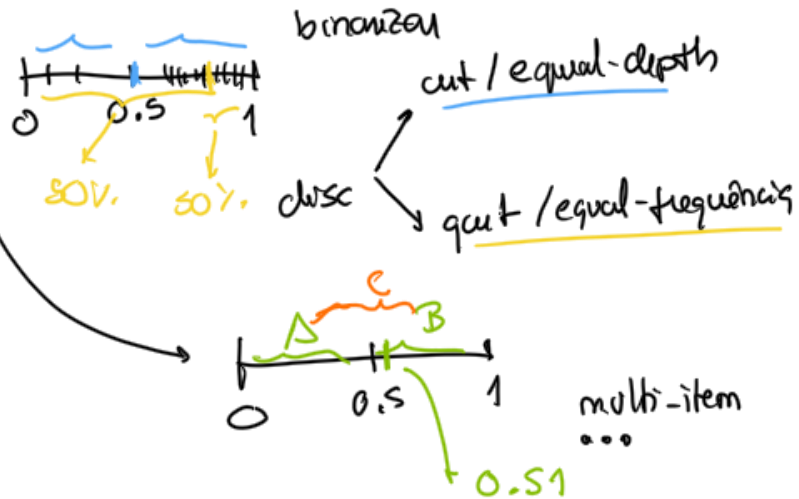
$A \cap B \Rightarrow C \cap D$

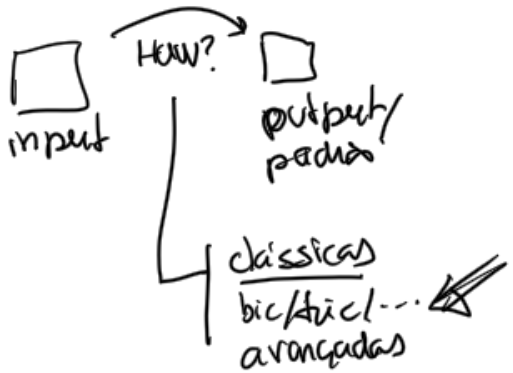
$\Rightarrow F$

(9)

características numéricas $y_1 < \theta_1 \wedge y_7 \geq \theta_7 \wedge y_{13} \leq \theta_{13} \Rightarrow \ldots$

↳ subspace clustering
↳ discretizar →





cat ⇒ cat
num ⇒ cat
cat ⇒ num
num ⇒ num

(10) multi-level patterns

↳ hierarquias de itens

(11)

clustering (regularidades "globais") → anomalias (irregularidades "globais")

padrão positivo (regularidades locais) inesp. frequente ⇒ padrão negativo (irregularidade locais) inesp. infrequentes/raras

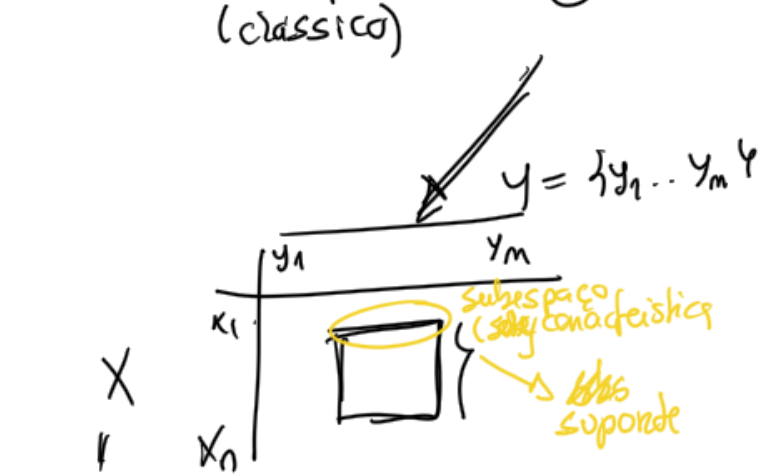
$P(\text{carne}) = 0.1$
$P(\text{tofu}) = 0.1$
$P_{null}(\text{carne} \cap \text{tofu}) = 0.1 \times 0.1 = 1\%$

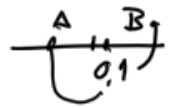
tofu ∩ carne $P(\text{carne} \cap \text{tofu}) = 0.001\%$

filme A ∩ filme B

① descoberta padrões (clássico) { ① numéricas → discretizar ② constantes

$y_7 > 0_7 \wedge 0 \leq y_6 \leq 0_6 \Rightarrow \dots$





	y_7	y_8	y_{19}
x_3	3	5	4
x_7	2	6	5

→ temos prods

	y_7	y_8
x_3	3	5.8
x_7	2.9	6

..

docs utilizadores

OP

constantes

{x1 .. xn}

biclusterings {B1 ... BK}

$B_j = (I, J)$ → $J \subseteq Y$, $I \subseteq X$

que satisfazem y_j { homogeneidade, sig. estatística, dissimilaridade

② homogeneidade → coerência → tipo: constantes; $c_j + \eta_{ij}$; aditivas; $c_j + \gamma_i + \eta_{ij}$

coerência → forma: $\eta_{ij} \in [-0.2, 0.2]$

coerência → ordem preservada

homogeneidade → qualidade

homogeneidade → estrutura

multiplicativas

	y_8	y_4	y_5
x_7	1	2	3
x_{10}	1	1.9	3.1
x_{11}	1	2.1	2.9

$Q = \{y_3 \approx 1, y_4 = 2, y_5 = 3\}$

$B_1 = (\{x_7, x_{10}, x_{11}\}, \{y_3, y_4, y_5\})$, $\{x_7, x_{10}, x_{11}\} = I$, $\{y_3, y_4, y_5\} = J$

	y_2	y_4	y_5
x_1	1	3	2
x_2	2	4	3
x_3	5	7	6

$\gamma_1 = 0$, $\gamma_2 = 1$, $\gamma_3 = 4$

$c_2\ c_4\ c_5$

$\langle 1, 3, 2 \rangle$

$c_j + \gamma_i + \eta_{ij}$

X

③

$\{B_1 .. B_K\}$ → homogeneidade ↔ coerência / qualidade

estrutura ↔ #bics / posicionamento

B1 | B2

	y_2	y_3	y_4
x_1	1	3	2
x_2	1	3	2
x_3	1	3	2

qualidade = 100%

qualidade $(B_2) = \frac{8}{9}$

$r_{ij} \notin [-\delta, \delta]$ $= 1 - \frac{1}{9} = \frac{8}{9}$

erros

avaliação

B=(I,J)

|I|

suporte/freq

sig. estatísticas *

↓ resp. frequente

$P(A \cap B)$

$P(2-\delta \leq y_3 \leq 2+\delta, 1-\delta \leq y_7 \leq 1+\delta)$

$B_j \Rightarrow C_j$

confiança

lift

// $y_3 \perp\!\!\!\perp y_7$

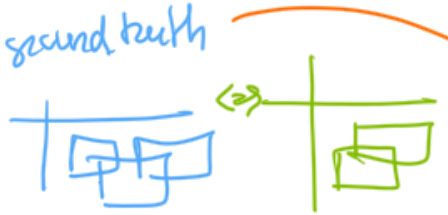
$P(2-\delta \leq y_3 \leq 2+\delta) P(1-\delta \leq y_7 \leq 1+\delta)$

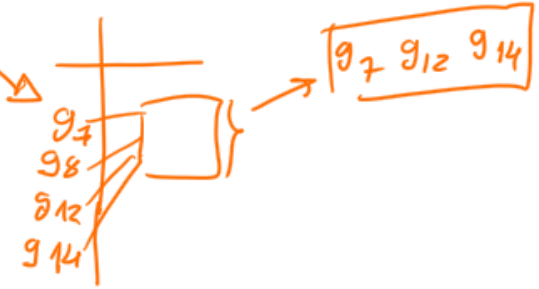
hierárquica

$\hat{n}$ supervisionado vs supervisionado

S → {dados reais → nha ground truth
dados sintéticos

clustering

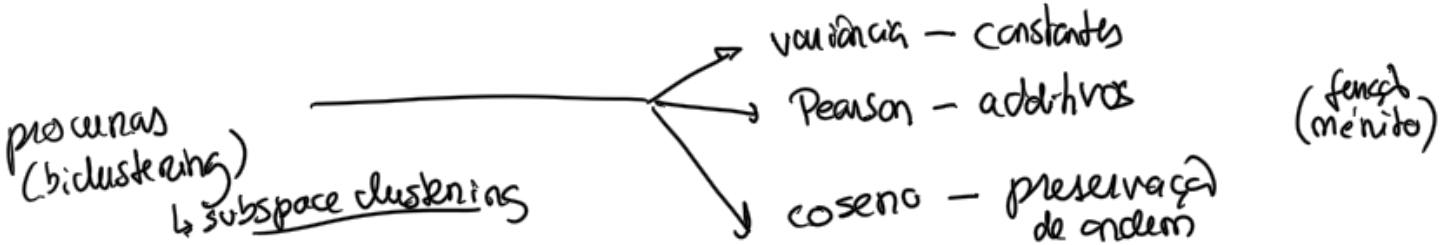


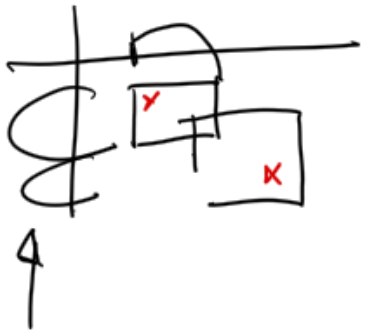


g_7 g_{12} g_{14}

* ≠P
≠N

8



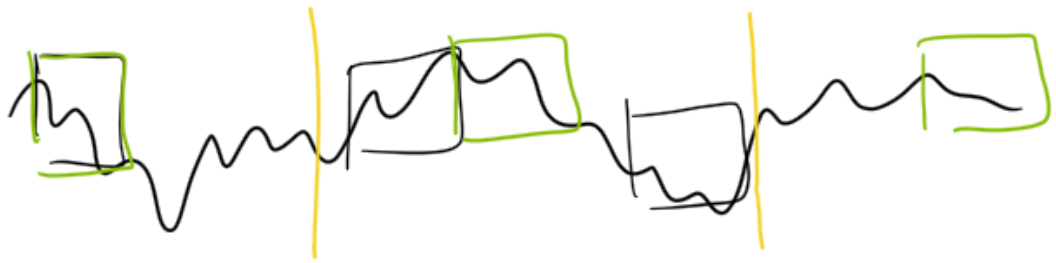


1	1	1
1	1	1
1	1	

→ variância

X_1	1	2	3
X_2	200	300	400

1	0.8	1
1.2	1	0.9



	y_3	y_4	class
x_1	B	D	X
x_2	A	C	X
x_3	A	C	X
x_4	A	D	Y

12) padrões simples / closed / maximal com min sup = 0.5

- simples: A, B, C, D, AC, AD, CD
- fechado: A, D, AC
- máximo: AC, D

13) R: $AC \Rightarrow X$ suporte confiança lift

$sup(AC \Rightarrow X) = 0.5$

$conf(AC \Rightarrow X) = \frac{sup(AC \Rightarrow X)}{sup(AC)} = \frac{2/4}{2/4} = 1$

$lift(AC \Rightarrow X) = \frac{conf(AC \Rightarrow X)}{P(X)} = \frac{1}{3/4} = 4/3 > 1$

p-value do padrão AC

$p(n \geq 2 \mid Bin \sim (N=4, P_{null}(AC) = 6/16)) = 0.15$

$P(A)P(C) = 3/4 \times 2/4$

$\alpha = 5\%$ / 1%

15) dados

	y_1	y_2	
x_1	-2	2	$y_1 = 0$
+2 x_2	3.5	4	$y_2 = 2$
+2 x_3	0	4	$y_3 = 2$
x_4	-2	2	$y_4 = 0$

	y_1	y_3	y_7
x_1	0.3	122	A
x_2	0.32		

16) aditivo calcula a qualidade 7/8

largest | constante bic / order-preserving bic

$\delta = 0$ qualidade = 100%

	y_6	y_8
x_3	1	1
x_5	1	2

$B = (1\, x_1, x_2, x_3, x_4\, 1\, y_1, y_2)$

$B = (1\, x_1, x_3, x_4\, 1\, y_1, y_2)$

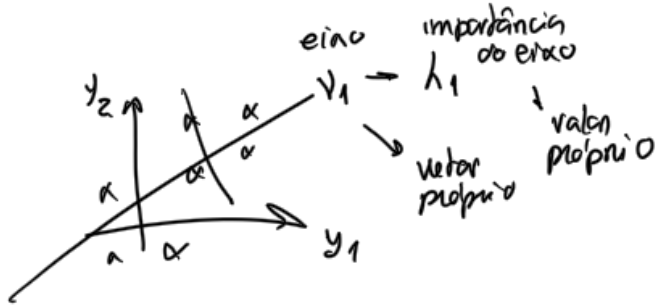
II PCA

numeric variables

	y_1	y_2
x_1	-2	2
x_2	3	4
x_3	0	4
x_4	-2	2

$v_1 = ?$ $\quad \lambda_1 = 0.3$

$v_2 = \begin{pmatrix} 0.9 \\ 0.4 \end{pmatrix} \quad \lambda_2 = 6.6$



9) % variabilidade explicada pela componente + informativa

$$\frac{\lambda_2}{\lambda_1+\lambda_2} = \frac{6.6}{6.6+0.3} \approx 95.6\%$$

10)

$$c_2 = 0.9\,y_1 + 0.4\,y_2$$

	y_1	y_2
x_1	-2	2
x_2	3	4
x_3	0	4
x_4	-2	2

X

	c_2
x_1	-1
x_2	4.3
x_3	1.6
x_4	-1

$$U = \begin{pmatrix} .9 \\ .4 \end{pmatrix}$$

$$X' = U^T X$$

$$= (.9\ .4) \begin{pmatrix} -2 & 3 \\ 2 & 4 \\ \cdots & \end{pmatrix}$$

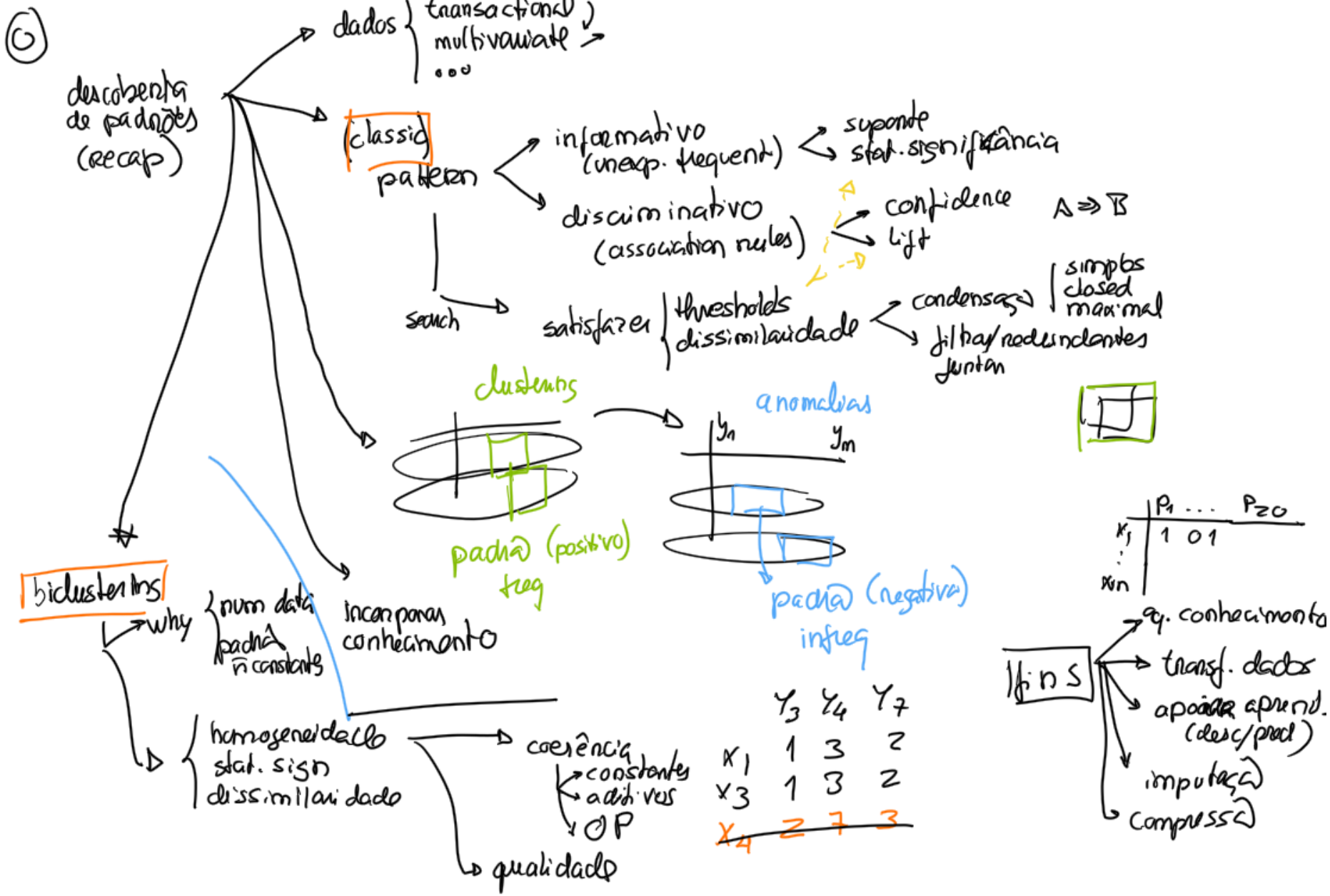
$$= (-1 \quad 4.3 \quad 1.6 \quad -1)$$

$$X = (U^T)^{-1} X'$$

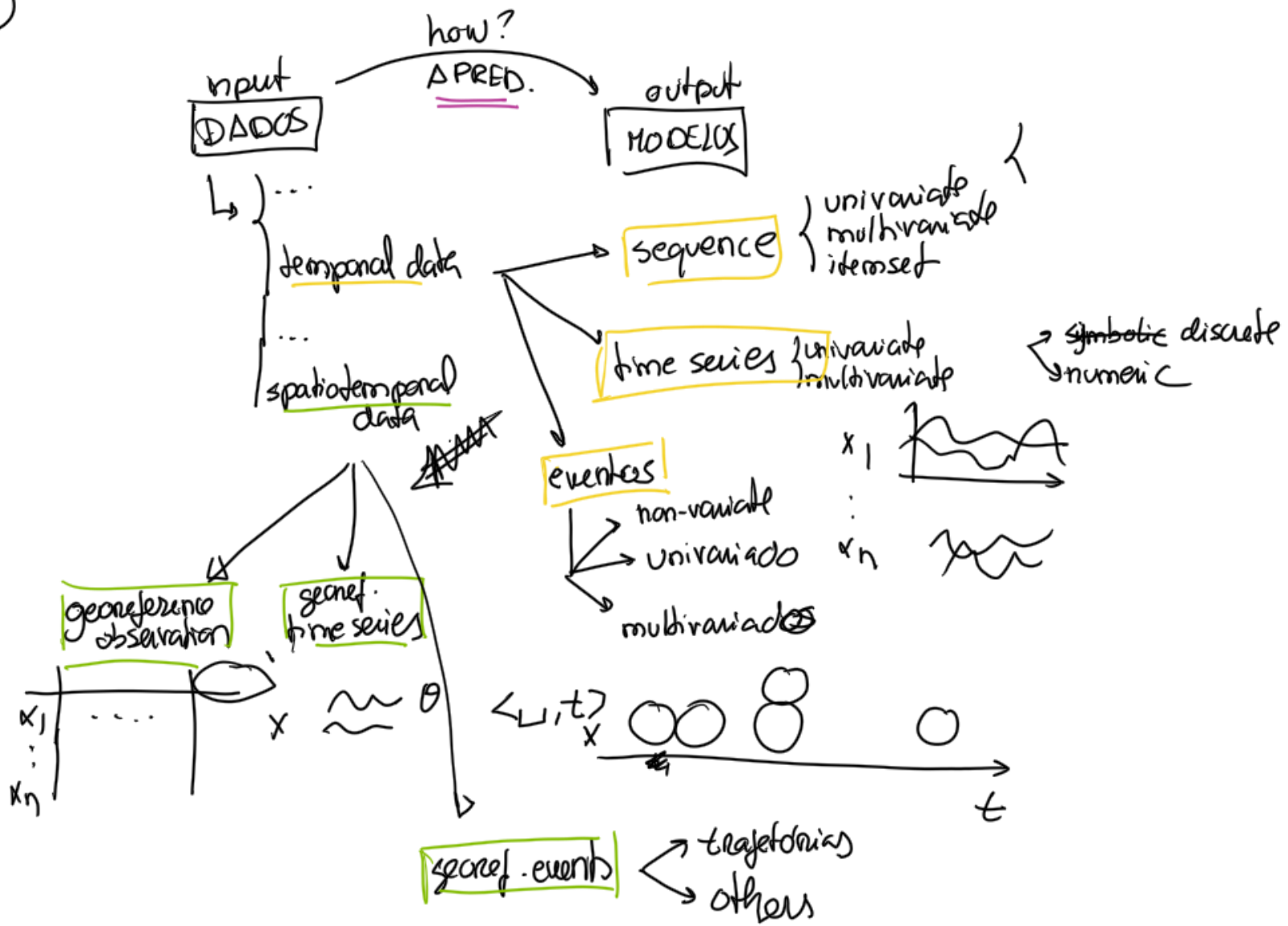
$$X = U X'$$

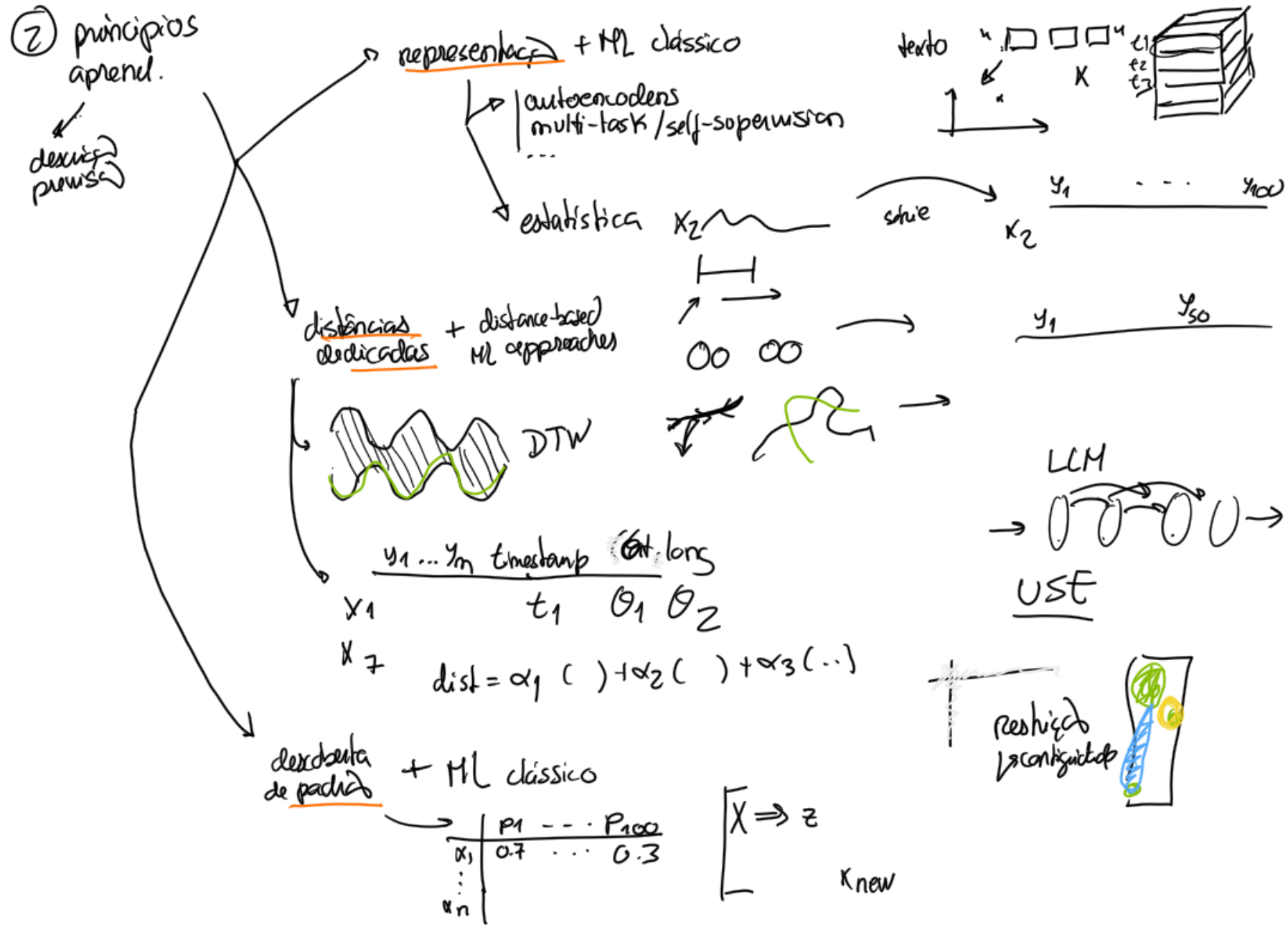
$$= \begin{pmatrix} .9 \\ .4 \end{pmatrix} (-1 \quad 4.3 \ \ldots)$$

$$= (\quad\quad)$$


6
descoberta de padrões (recap)
dados
transactional
multivariate
...
(classic) pattern
informativo (unexp. frequent)
suporte
stat. significância
discriminativo (association rules)
confidence
lift
A ⇒ B
search
satisfazer
thresholds
dissimilaridade
condensação
simples
closed
maximal
filhas/redundantes juntas
clusters
anomalias
padrão (positivo)
freq
padrão (negativa)
infreq
biclusterings
why
num data
padrão
não constante
incorporar conhecimento
homogeneidade
stat. sign
dissimilaridade
coerência
constantes
aditivos
OP
qualidade
Y3 Y4 Y7
X1 1 3 2
X3 1 3 2
X4 2 7 3
P1 ... P20
X1 1 0 1
Xn
fins
q. conhecimento
transf. dados
apoiar aprend. (desc/prod)
imputação
compressão

1


how?
APRED.
input
DADOS
output
MODELOS
temporal data
spatiotemporal data
sequence
univariate
multivariate
itemset
time series
univariate
multivariate
symbolic discrete
numeric
events
non-variate
univariado
multivariados
georeferenced observation
georef. time series
georef. events
trajectorias
others
t

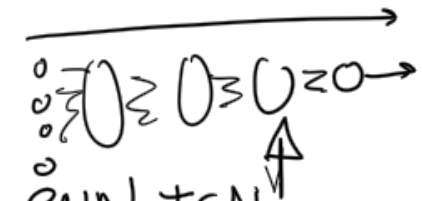

2 princípios aprend.
descrição premissa
representação + ML clássico
autoencoders
multi-task / self-supervision
...
estatística
X_2
série
y_1 ... y_{100}
X_2
distâncias dedicadas + distance-based ML approaches
y_1 y_{50}
DTW
$y_1 \dots y_m$ timestamp (int, long)
t_1 θ_1 θ_2
X_1
X_7
$dist = \alpha_1 (\ ) + \alpha_2 (\ ) + \alpha_3 (..)$
texto
K
LCM
USE
redução dimensionalidade
descoberta de padrões + ML clássico
P1 ... P100
x_1 0.7 ... 0.3
x_n
$X \Rightarrow z$
Knew

(2)

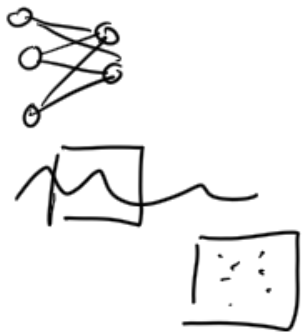
cont.

principios aprend.

deep learning (NNs) end-to-end



- tempo → CNN, RNN, TCN
- espaço → CNN, GNN
- ...
- HMMs DBNs →



prompting (prompt engineering)

→ zero-shot / few-shot

⊖ numerical precision
⊖ text vs data attention
⊖ hallucination
⊖ sensitive prompt assumptions

→

⊕ be specific
⊕ scenario based
⊕ uncertainty
⊕ resolution scope

LLM → (supervised) fine tuning

input-output

3 Padrões

- sequências (ex. itemset seq) → co-ocorrência precedência
- shapes
  - single → motifs analysis { univ → unid. multiv → multid.
  - data → x_1, x_2, x_3 → t_1 … t_p × x_1 … x_n
    - univ → biclustering { contiguidade temporal, tolerar desalinhamento
    - multiv → triclustering
- eventos
  - episodes (A → B, C → B) partial orders
  - duração (A, B, B, C) → t
  - t_p × t_1 … t_m × x_1 … x_n → p_1 … p_e × x_1 … x_n

$x_1 = \langle \binom{a}{b} (c) \binom{a}{d} \binom{a}{c} \rangle$

$x_2 = \langle \binom{b}{d} (d) \binom{a}{b} (d) (e) \rangle$

padrão sequencial → $\binom{a}{b}(c)$ suporte sig. estatísticas

$(a)(c)(d)$ → $P(a)P(c|a)P(d|c)$

padrão seq ⇒ alvo conf lift

	t_1	t_2	t_3	t_4	t_5
x_1	1	3	2	4	1
x_2	0	1	3	2	4

③ cont.

padrões → espaço-temporais

$A \xrightarrow{\Delta t} B$

$A_\theta \Rightarrow B_\theta$

baricentro

medoide

cluster

componentes

④

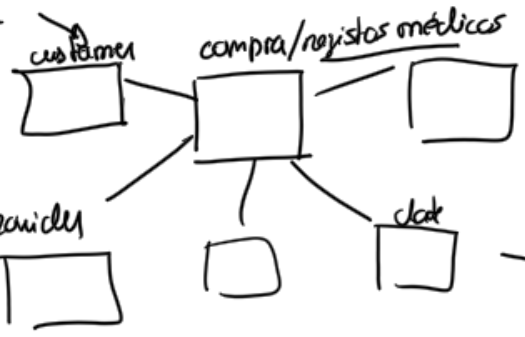
Relacional

multi-dimensional

ILP

$has(x, B) \wedge likes(x, C) \Rightarrow prefer(B, C)$

atomset



hiclustering

...

desc

pred

$sex(x, F) \wedge ordu(x, y) \wedge value(y) > 1000 \Rightarrow credit = yes$

Exam 2022

	y_1	y_2	y_3	y_4	dens	cluster
x_1	-2	2	B	D	X	c_1
x_2	3	4	A	C	X	c_2
x_3	0	4	A	C	Y	c_1
x_4	-2	2	B	D	Y	c_1

$$d(x_A, x_B) = \text{Manhattan}_{num}(x_A, x_B) + \text{Hamming}_{cat}(x_A, x_B)$$

I clustering

1)

	x_1	x_2	x_3	x_4
x_1	0	9	?	?
x_2		0	3	8
x_3			0	5
x_4				0

6

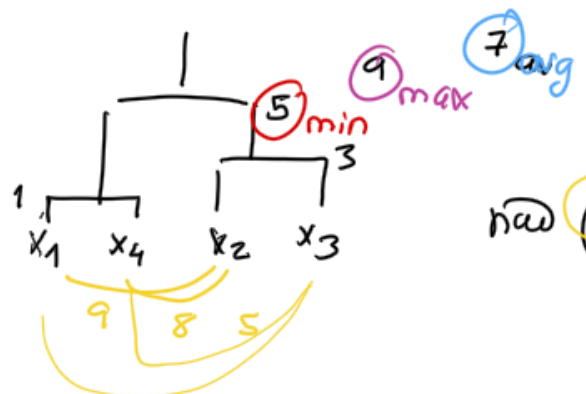
1

1) $d(x_1, x_3) = |-2-0| + |2-4| + 1 + 1$

$= 2 + 2 + 1 + 1 = 6$

$d(x_1, x_4) = 0 + 0 + 1 + 0 = 1$

2)



3) $\bar{c}_1 = x_1 \quad c_1 = \{x_1, x_4\}$

$\bar{c}_2 = x_4 \quad c_2 = \{x_4, x_2, x_3\}$

novo $(\{x_1, x_4\}, \{x_2, x_3\})$ aglomerativo

$\neq (\{x_2\}, \{x_1, x_3, x_4\})$ divisivo

$d_{média}(x_3 | c_2) = \frac{d(x_3, x_2) + d(x_3, x_4)}{2} = \frac{3+5}{2} = 4$ 6

$d_{média}(x_2 | c_2) = 5.5$

$d_{média}(x_4 | c_2) = 6.5$

após iteração

$\bar{c}_1 = x_1$

$\bar{c}_2 = x_3$

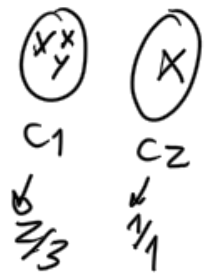
novos medóides

Exam 2022

I 4) silhouette $s(x_4) = 1 - \frac{a(x_4)}{b(x_4)} = 1 - \frac{\frac{1+5}{2}}{8} = 5/8$

	x_1	x_2	x_3	x_4	cluster	class
x_1	0			1	c_1	x
x_2		0		8	c_2	x
x_3			0	5	c_1	x
x_4				0	c_1	y

5) purezza $(C, G) = \frac{1}{4}(2+1) = 3/4$



Autoencoder

$W^{[1]} = \begin{pmatrix} 1 & 1 & -1 & 0 \\ 1 & -1 & 0 & -1 \end{pmatrix}$ $b^{[1]} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$W^{[2]} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}$ $b^{[2]} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$

$\phi(x) = ReLU(x)$

$\max(0, x)$

$x_{new} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = x^{[0]}$

MAE $= \frac{0+0+0+0}{4}$

RMSE $= \sqrt{\frac{1}{4}(0^2+\dots)} = 0$

g) embedding $z_{new} = x^{[1]}$

$= \phi(W^{[1]} x^{[0]} + b^{[1]})$

$= \phi\left(\begin{pmatrix} 1 & 1 & -1 & 0 \\ 1 & -1 & 0 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right)$

$= \phi\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) = \phi\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

b) $\hat{x}_{new} = x^{[2]} = \phi\left(\begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}\right) = \phi\left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$