Dx

"Problem statement" ☑

eg: sales data

solve

Discount, Blul → ?

labbel

ABL

Day	Discout	Blh1	Volume
1	10%	F	300
2	9%	F	302
3	1%	T	280
¦	¦	¦	

X ← input

Y ← output

"Historical"

(8%, F)

300

predictn
250

(8%, F)   Real

Esam

Book
↑
data

?   A
A
10K

training methods

(X, Y)

D_x —

X

X →↑ Real ↑

Historical data {

Discout	B\h\l	Volume
—	—	—
—	—	—
—	—	—
—	—	—
—	—	—

[ ] ? 300
         ?

? (box) ↑y→
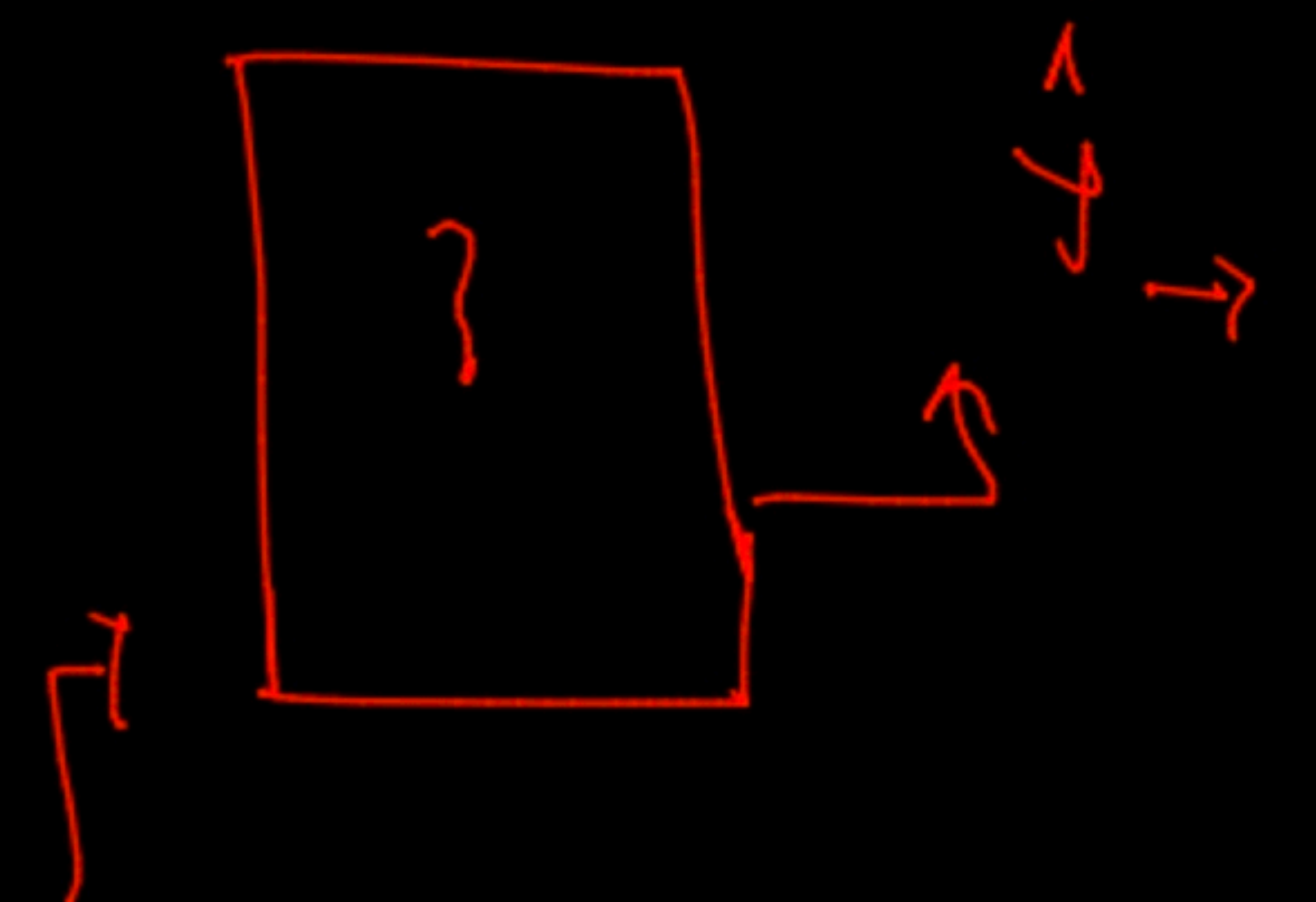
$$D_x (x, y)$$

ML

* No labelled
$$D_x (x) \uparrow^{y}$$

eg.

customer-segmentation
⇕
iphone 17 X

"Supervised ML"

Unsupervised ML

$Y \in R$

"Classification"

Regression
⇕
1) Linear Regression

History

cid	OS	han	digital
C1	-	-	T
C2	-	-	
C3	-	-	1
C4	-	-	
C5	-	-	1
C6	-	-	
:	:	:	:
C10K	-	-	

1) KNN
2) Logistic Regression
3) SVM (
4) Random Forest
5) Decision Tree
6) }

{ ←

$\downarrow$
$\downarrow_x$

sklearn e|ML

theory → imple

Iphone

?

X
↑

Y

eg.

	cid	income	cs	status
	1	14	750	Yes
	2	15	650	Yes
	3	6	350	No

X → income, cs
Y → status

Historical

DX

↑

labbeled

SML

ML

→ (income, cs)

↑ (20, 750)

Y → Yes
   → NO

∈ [Yes, No]

→ classification

qs
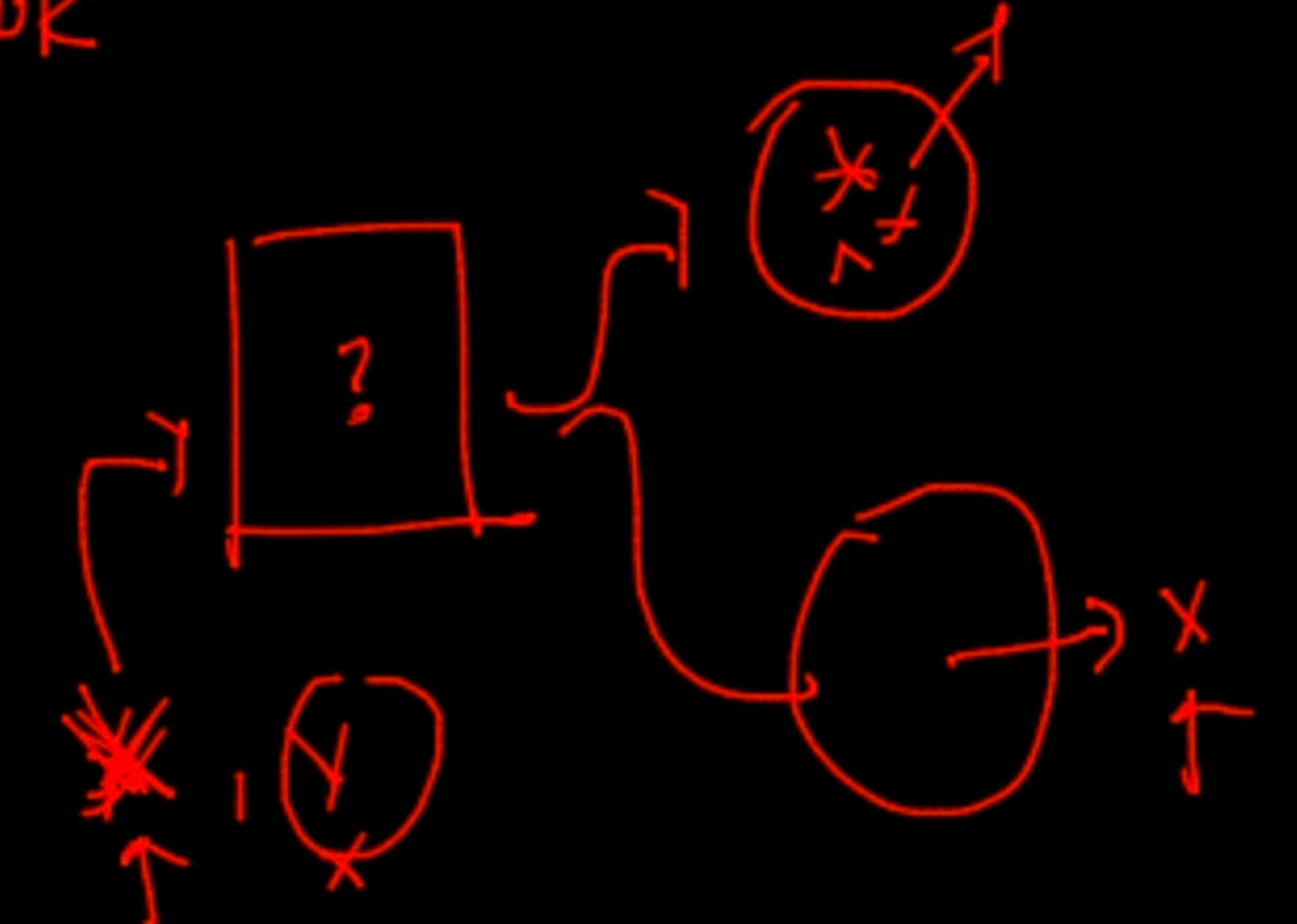
Algorithm
↑

KNN (k-Nearest Neighbours) → simple →
‾‾‾   ↑
↑

Exam
↑

'Book'

Q - A
Q₁ A
Q₂ A
⋮
⋮
Q_LUK A

training

testing

→ production

Exam
‾‾‾
→ judge → score
→ 'A'   ↑
=

↑
a
allowed
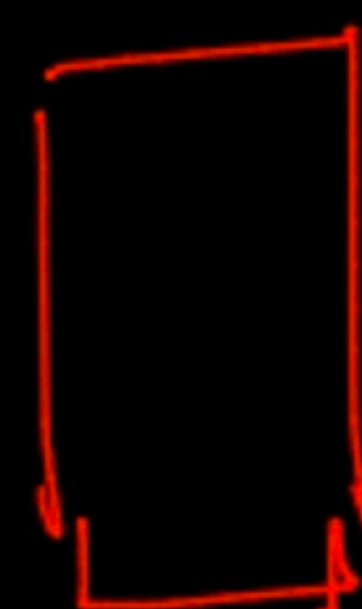
Idx

cid   Income   cs   status

           15   750 ↝   yes

②

                                    → pass

                                         ⊕
                                         1   → very well

10k   —   —   —

'testing'

▷ prelim

splitting Before trning

$D_X(X,Y)$

Random split

$D_{Train}(X_{train}, Y_{train}) \rightarrow 80\%$ NO

$D_{Test}(Y_{test}, Y_{test}) \rightarrow 20\%$

Yes

eg.

D_X
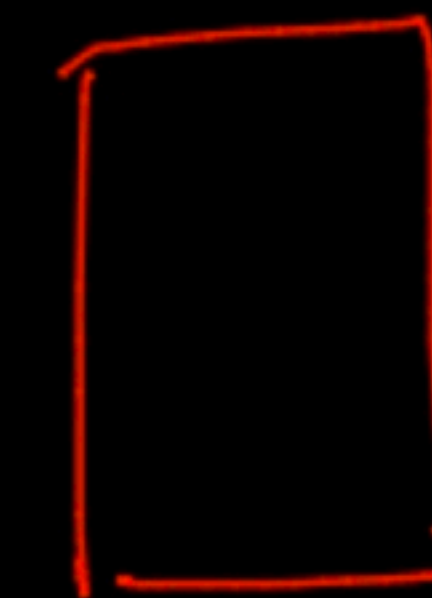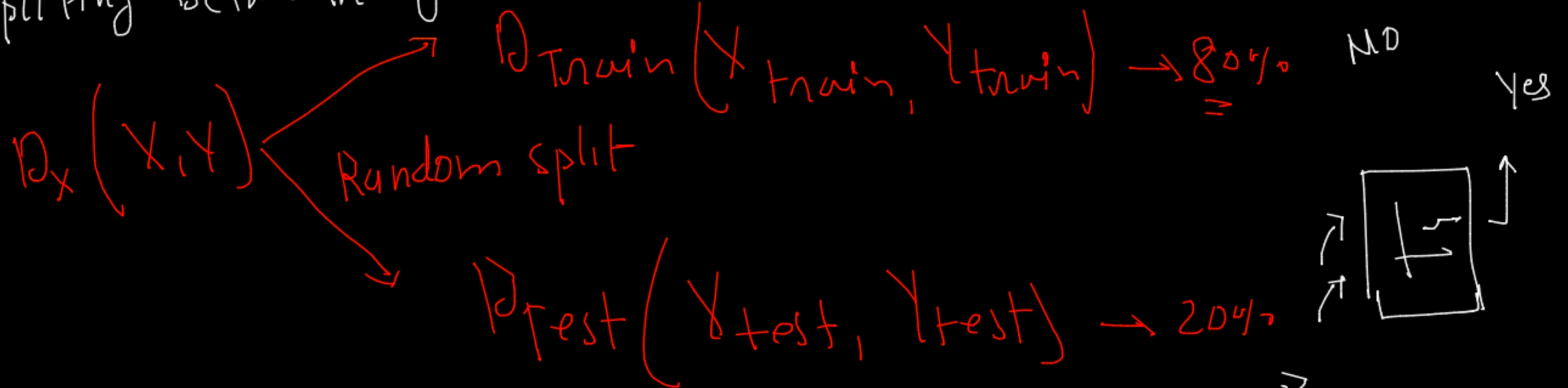
i	B	Y
14	750	Y
13	600	Y
14	350	N
11	450	N
16	75	N
12	600	Yes

Random spliting

14	75'	Y
11	45'	N
14	35'	N

$\rightarrow$ to train

13	150	N	Y
12	65'	Yes	N

$\rightarrow$ testing

$\frac{1}{2} \rightarrow 50\% = \frac{1}{2}$

X

Dx ↑

classify 6 5 4 3 2

KNN

cid | Income | CS | Status
1 | ·15 | 750 | Yes → 4.02
2 | 17 | 760 | No → 4.03
3 | 18 | 780 | Yes → 3.93 training
4 | 13 | 550 | No → 2.66
5 | 12 | 650 | Yes → 1.04 → sort
6 | 6 | 750 | Yes → 0.94 Historical
7 | 4 | 350 | No → Yes
8 | 9 | 420 | No → No
9 | 17 | 780 | Yes → Yes

of testing ↑

'(4,350)' → ? Yes

Yes / NO

"memory"
↑
U

2 - Yes
1 - NO

KNN

spliting

KNN
BNN
↑ ↑ Hyper para

Dx { Dtrain
     Dtest

2/3   66 to 7

CS

income