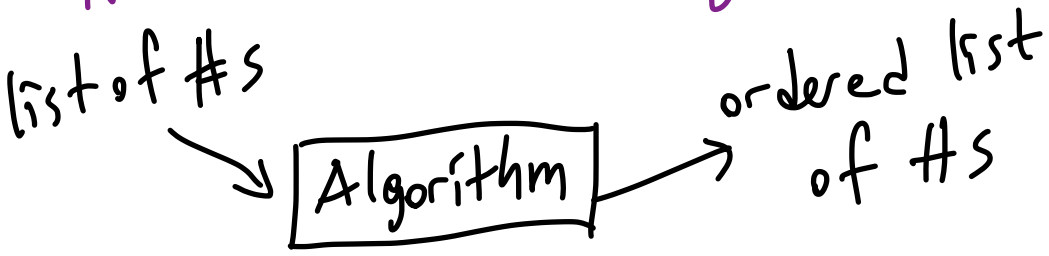
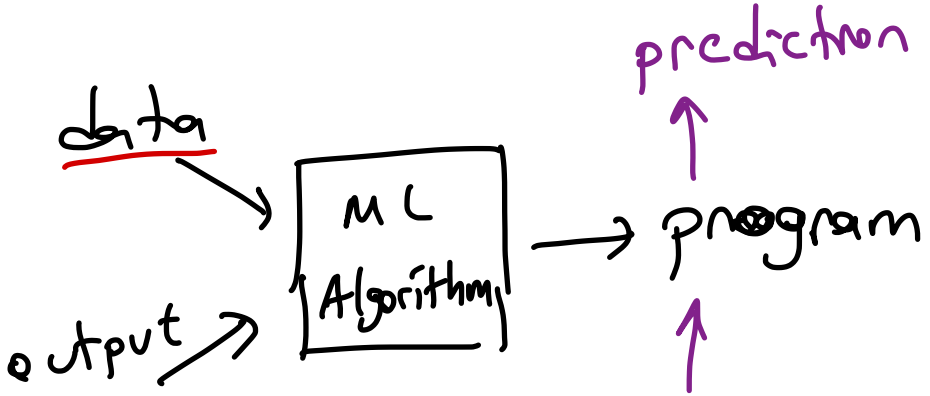
Algorithm vs ML Algorithm

- sorting
- searching
- ...

- given a list of #s
- task: to order these #s from smallest to largest

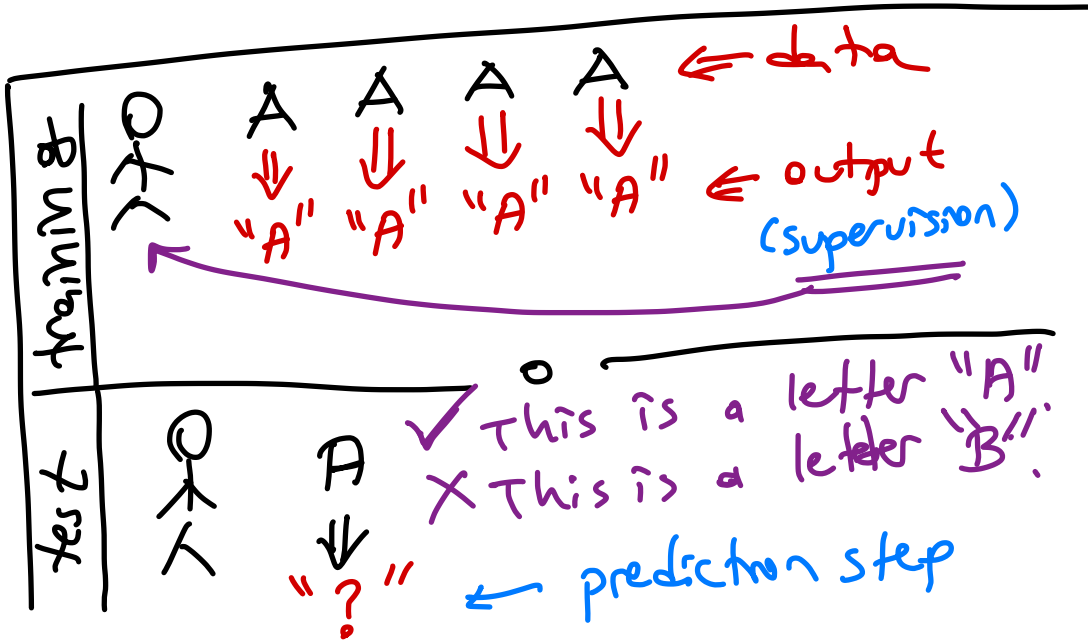


- fixed set of rules
- deterministic



- subjectivity (example data, past experience)

- test data
- unseen data
- out-of-sample data

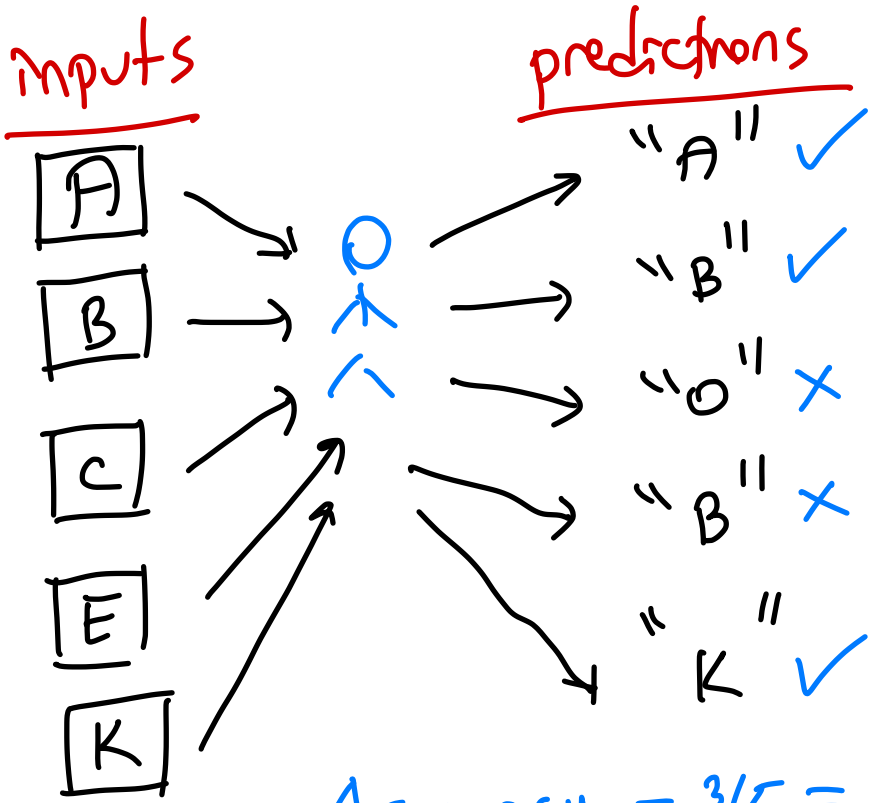


Machine Learning: programming computers to optimize learning a performance criterion using example data or past experience.

loss/error function

training set

↳ depends on the application



Accuracy = 3/5 = 60%

Error = 2/5 = 40%

error		Prediction	Truth
4°C	Tuesday	10°C	14°C
4°C	Wednesday	12°C	8°C
4°C	Thursday	8°C	12°C
1°C	Friday	11°C	12°C

absolute error

$|truth - predicted|$

$|y_i - \hat{y}_i|$

truth — y_i

predicted — $\hat{y}_i$

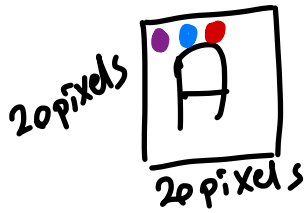
Supervised Learning:

$$\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N}$$

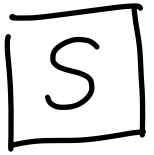
- $\mathcal{X}$ → training set
- x_i → i^{th} training data point
- y_i → i^{th} label, i^{th} target output
- N → # of data points
- i → index for data points

$$\mathcal{X} = \{(x_1, y_1), (x_2, y_2), \dots\dots (x_N, y_N)\}$$

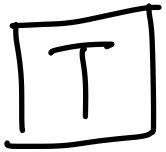
Classification:



x_1 "A" y_1 $\quad x_1 = \begin{bmatrix} \\ \\ \\ \end{bmatrix}_{400\times 1}$ $\quad y_1 = [A]$



x_2 "S" y_2 $\quad x_2 = \begin{bmatrix} \\ \\ \end{bmatrix}_{400\times 1}$ $\quad y_2 = [S]$



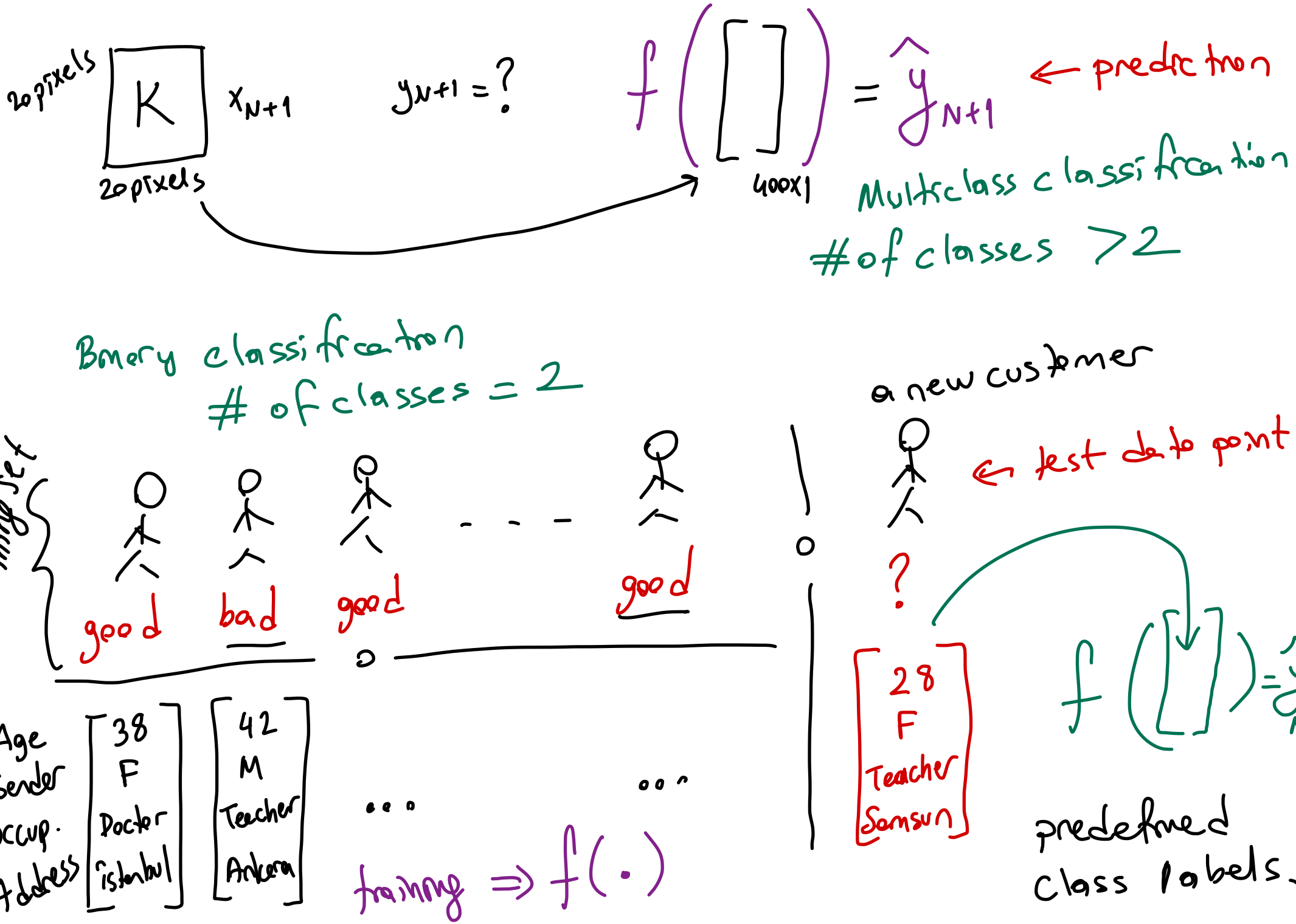
x_3 "T" y_3 $\quad x_3 = \begin{bmatrix} \\ \\ \end{bmatrix}_{400\times 1}$ $\quad y_3 = [T]$

$$\mathcal{X} = \left\{ \left(\begin{bmatrix} \\ \\ \\ \end{bmatrix}_{400\times 1}, \square \right) \right\}_{i=1}^{N}$$

$x_i \in \mathbb{R}^{400}$

$y_i \in \{A, B, C, \dots, Z\}$

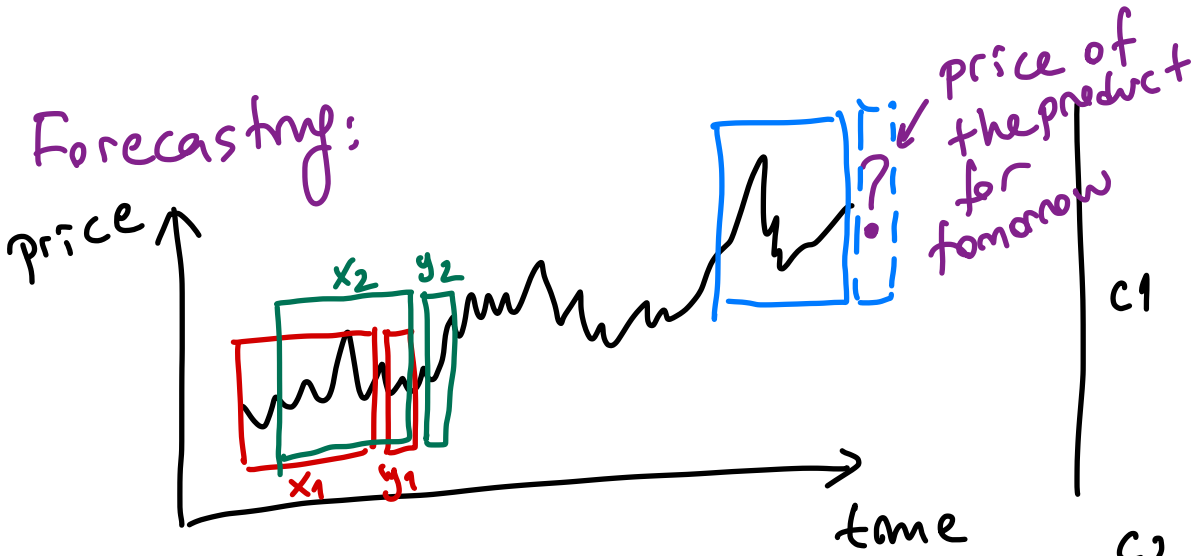
learning an $f(\cdot)$ function


20pixels
K
20pixels
x_{N+1}
$y_{N+1} = ?$
$f\left(\begin{bmatrix}\\\\\end{bmatrix}\right) = \hat{y}_{N+1}$
400x1
← prediction
Multiclass classification
# of classes > 2
Binary classification
# of classes = 2
a new customer
← test data point
training set
good
bad
good
good
?
Age
Gender
occup.
Address
38
F
Doctor
Istanbul
42
M
Teacher
Ankara
28
F
Teacher
Samsun
training ⇒ $f(\cdot)$
$f\left(\begin{bmatrix}\\\\\end{bmatrix}\right) = \hat{y}_{N+1}$
predefined
class labels.

Regression:

$$\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N} \qquad x_i \in \mathbb{R}^D \qquad y_i \in \mathbb{R}$$

Forecasting:

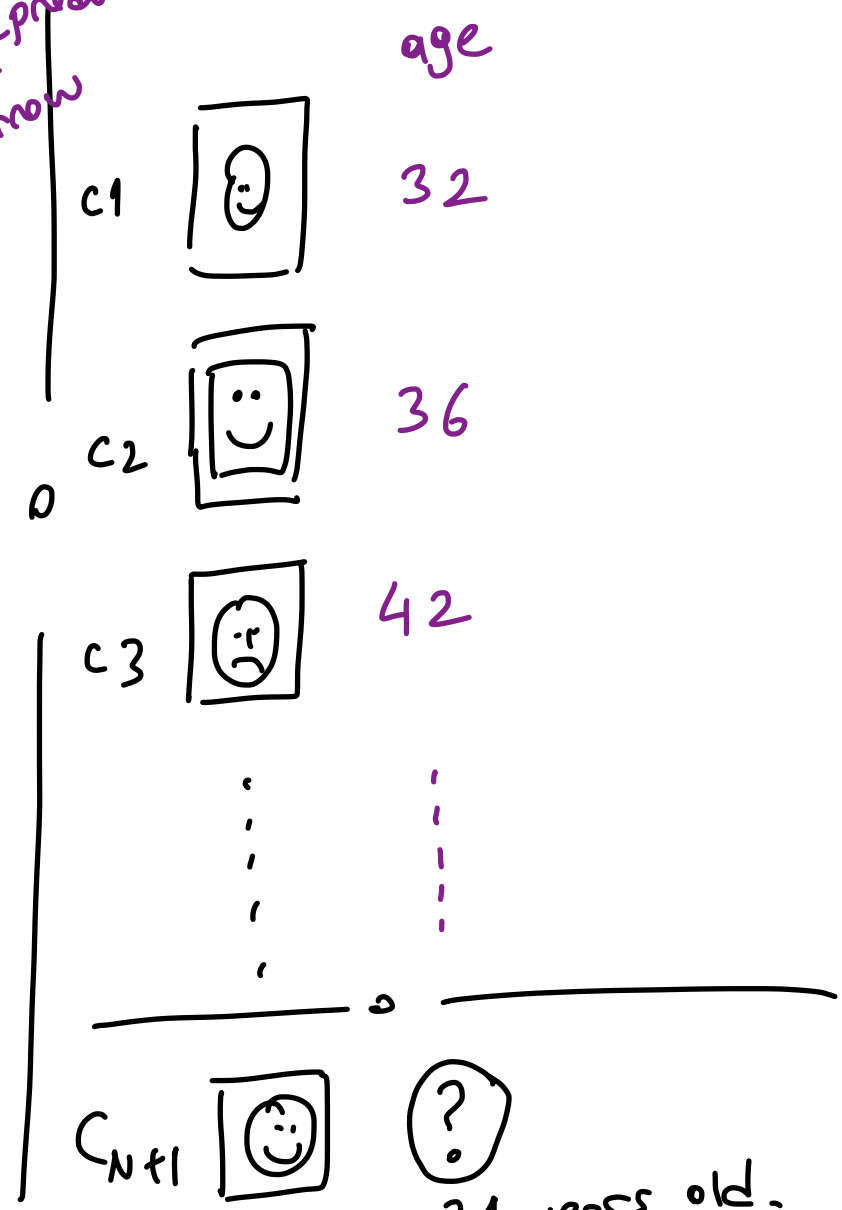


given the price for last 10 days, predict the price for tomorrow

$$x_1 = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{10} \end{bmatrix} \qquad y_1 = p_{11}$$

$$x_2 = \begin{bmatrix} p_2 \\ p_3 \\ \vdots \\ p_{11} \end{bmatrix} \qquad y_2 = p_{12}$$

$\vdots \qquad \vdots$



Unsupervised Learning:

Clustering $\quad \chi = \{ x_i \}_{i=1}^{N}$

NO CLASS LABEL !