INTRODUCTION TO MACHINE LEARNING COMPSCI 4ML3

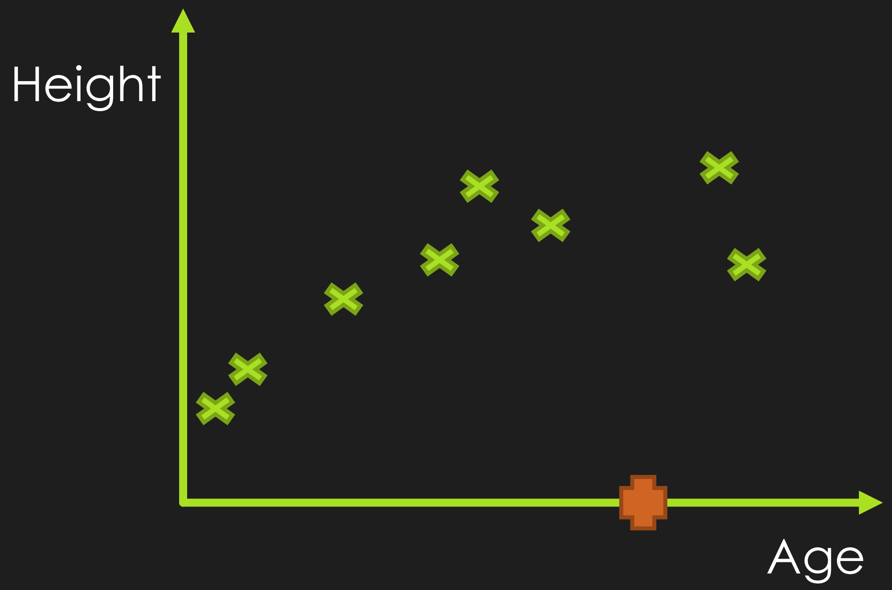
Lecture 1

Hassan Ashtiani

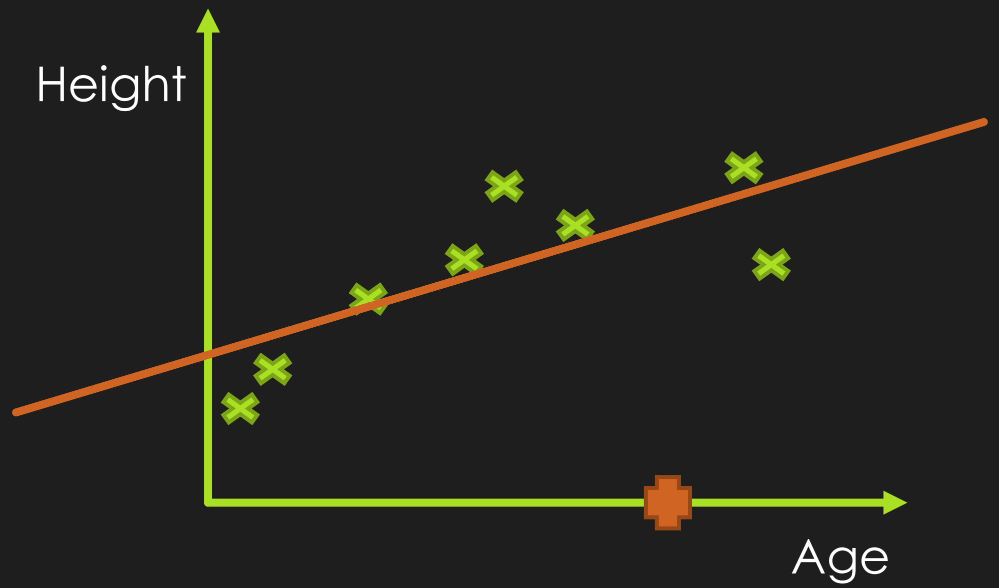
CURVE-FITTING

- Predict height of a person given her/his age?
- Collect a set of "data points"
- Represent data point i by (x^i, y^i)
  - E.g., if the individual i is 25 years old and is 175cm tall then we can write $(x^i, y^i) = (25, 175)$
- We have collected n data points, $S = \{(x^i, y^i)\}_{i=1}^n$
  - Given a new x, "predict" its y?

CURVE-FITTING

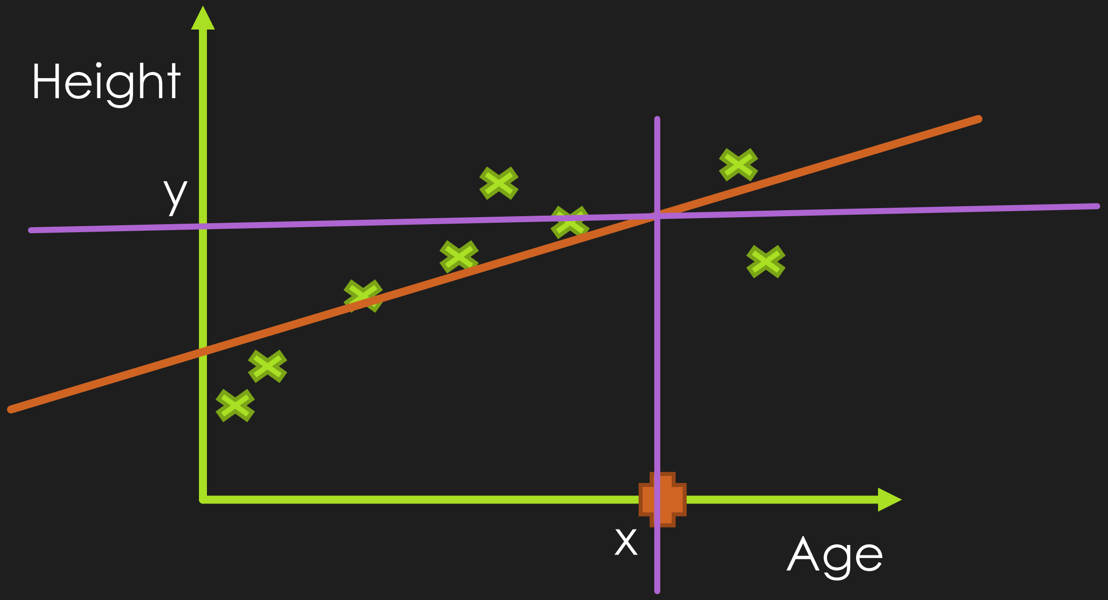


LINEAR CURVE-FITTING



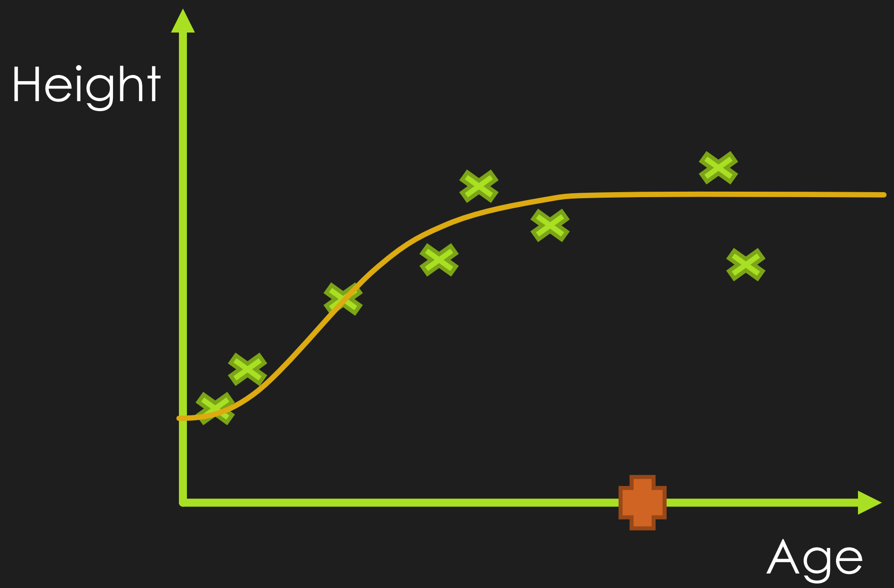
LINEAR CURVE-FITTING

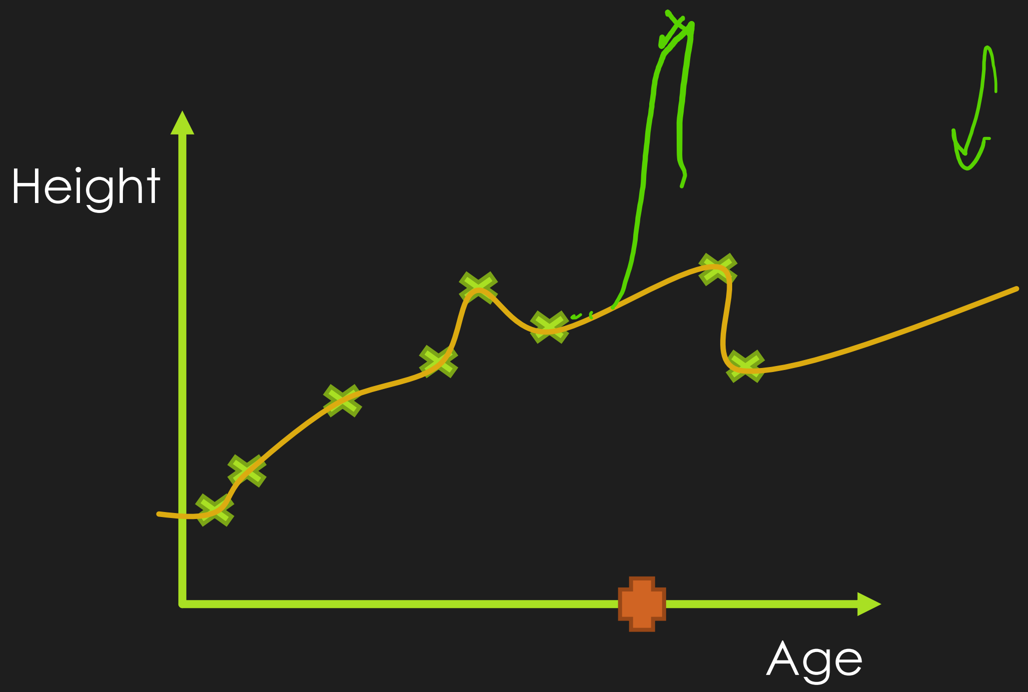
- x IS CALLED AN INPUT, y IS CALLED AN OUTPUT/RESPONSE



NON-LINEAR CURVE-FITTING

- Which one is better?



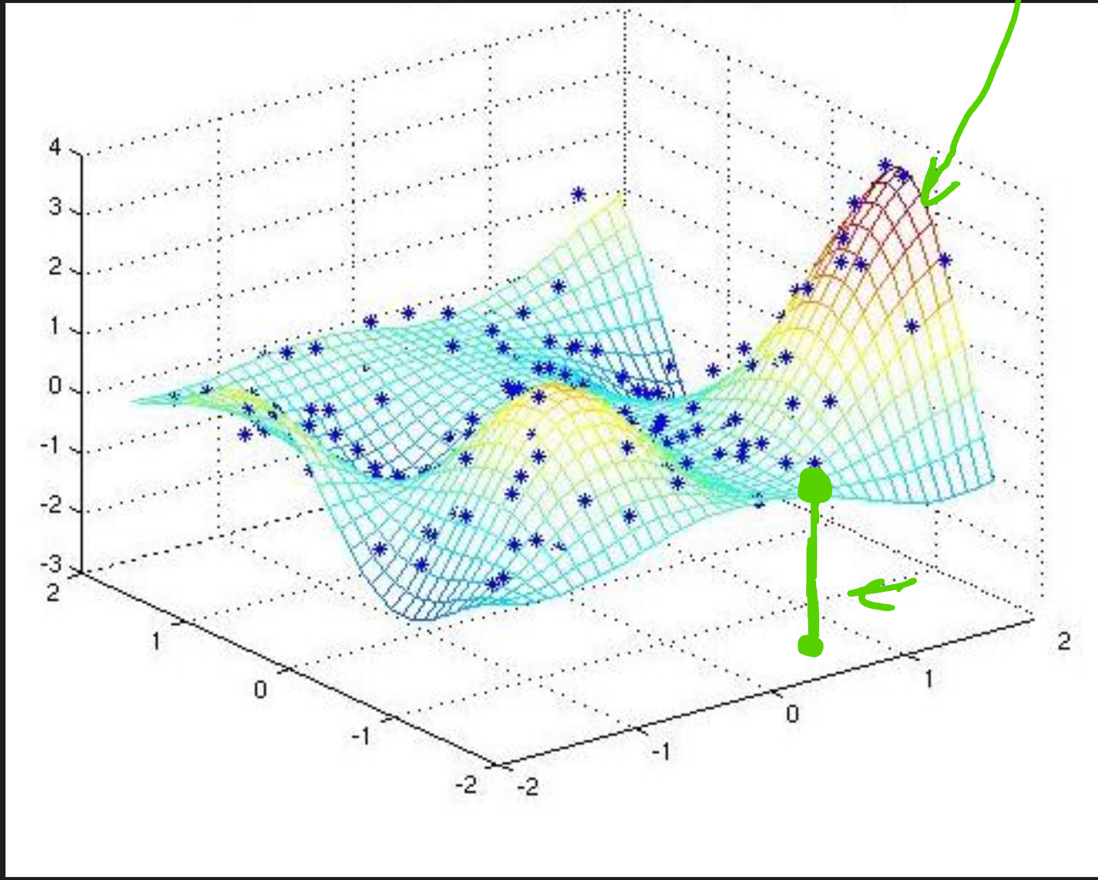


- Linear vs non-linear?

MULTIDIMENSIONAL CURVE-FITTING

$\hat{y} = f(x)$

- x AND/OR y COULD BE MULTI-DIMENSIONAL
- FOR EXAMPLE,
  PREDICT THE
  HEIGHT
  BASED ON THE
  AGE **AND** WEIGHT
- E.G., IN THE RIGHT PICTURE,
  $x \in \mathbb{R}^2, y \in \mathbb{R}$

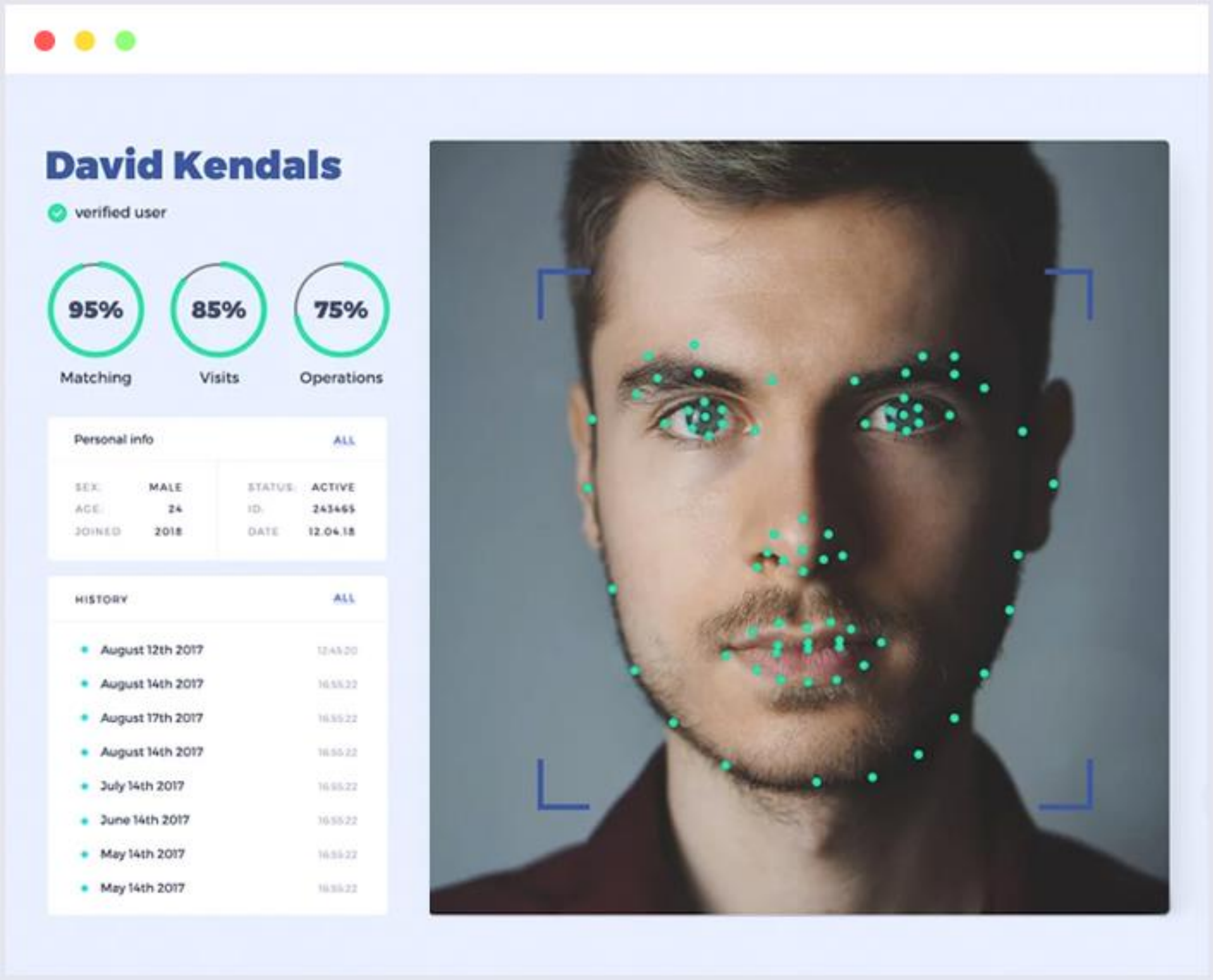


CURVE-FITTING IS EVERYWHERE

- Fitting a curve enables interpolation and extrapolation
- This is a type of supervised learning/prediction
  - Prediction, because we predict y given x
  - Supervised, because $\{(x^i, y^i)\}_{i=1}^n$ is given

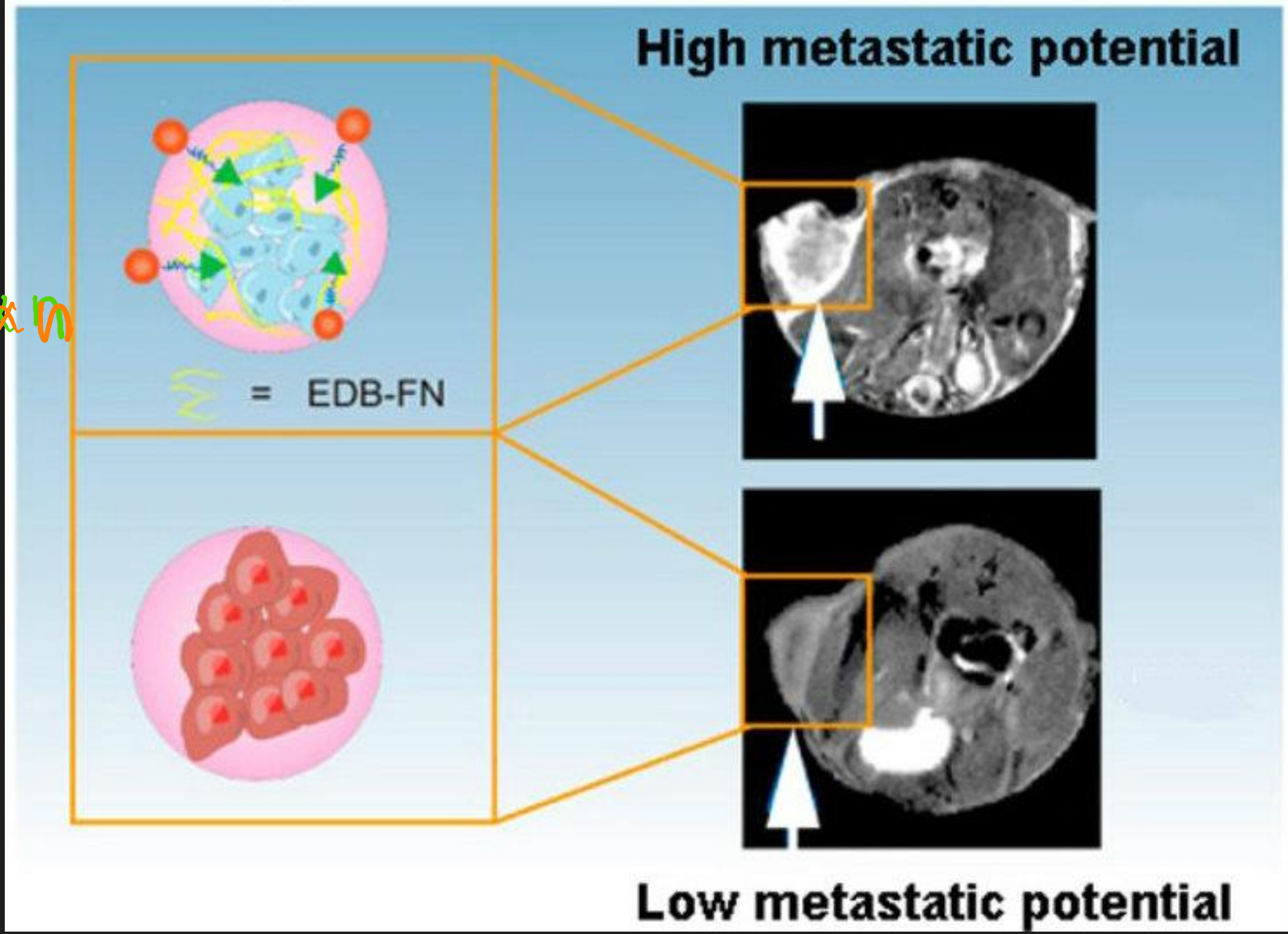
PREDICTION IS EVERYWHERE

- Face recognition
- x: image $m \times n \times 3$
  - $x \in \mathbb{R}^{m \times n \times 3}$
- y: identity
  - $y \in \{1, 2, 3, \dots, K\}$



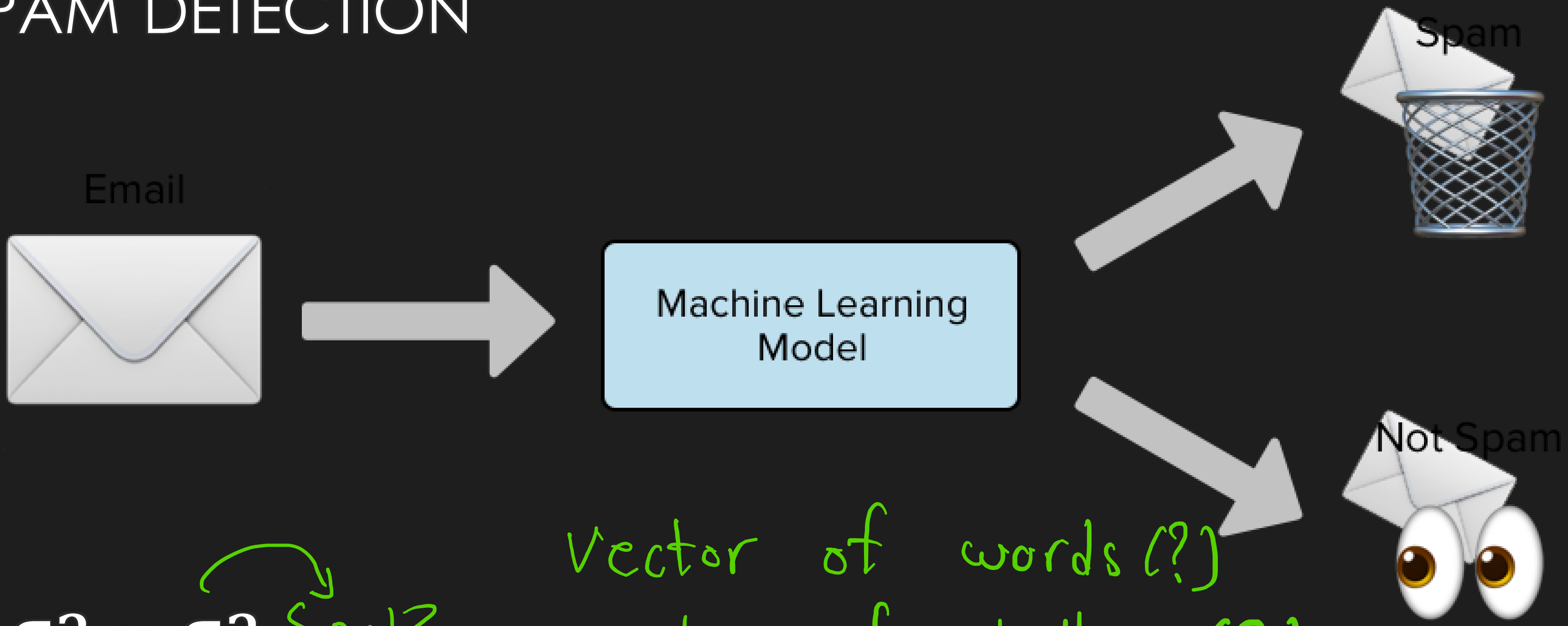
PREDICTION IS EVERYWHERE

- BIOMEDICAL IMAGING
- x: MRI IMAGE
  - $x \in \{0, 1, \dots, 255\}^{m \times n}$
- y: CANCEROUS?
  - $y \in \{0, 1\}$



PREDICTION IS EVERYWHERE

- Spam detection



Vector of words (?)
vector of letters (?)
sender, images, ...

- $x \in ?\ y \in ?$ $\{0,1\}$

PREDICTION IS EVERYWHERE

- OIL/STOCK PRICE PREDICTION
- E.G., GIVEN PRICE
  FOR $t = 1,2, \dots, 1000$

  PREDICT PRICE
  FOR $t = 1001$

- $x \in ?$ $\mathbb{R}^{1000}$
- $y \in ?$

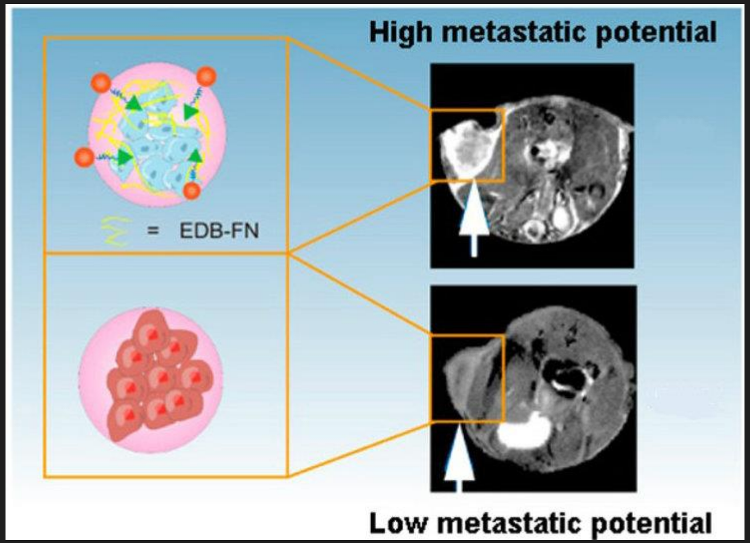
PREDICTION IS EVERYWHERE

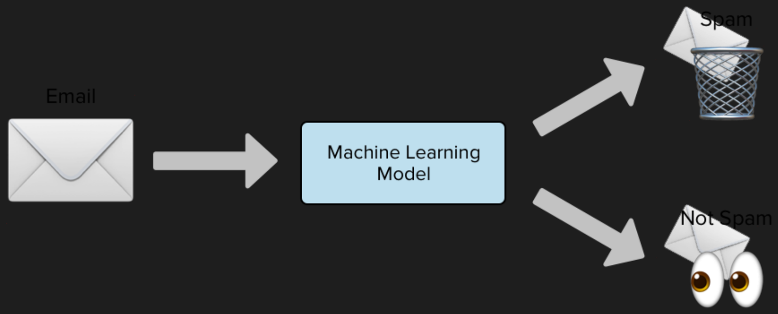
- Translating French text to English text
  - Input? Output? Sequence to sequence
- Speech recognition (Input? Output?)
- Text to speech (Input? Output?)
- Netflix recommendations (Input? Output?)
- Is everything in AI just prediction?!
  - Not exactly (e.g., prediction vs control)

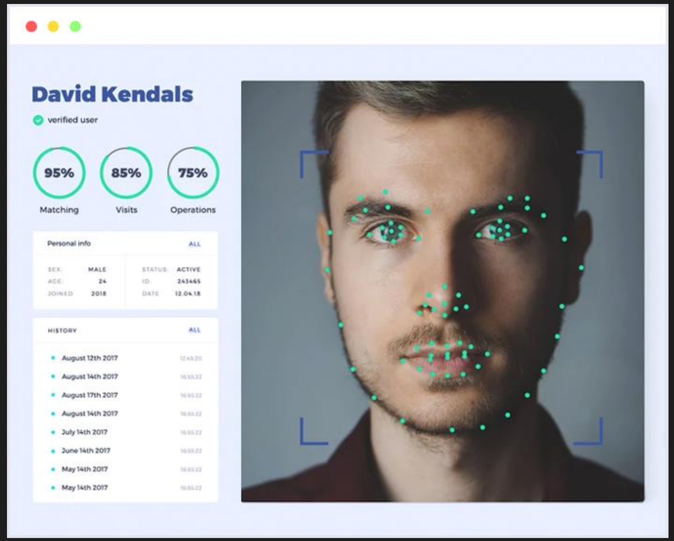
PREDICTION IS EVERYWHERE

PREDICTION IS EVERYWHERE

- AND … PREDICTION METHODS CAN BE QUITE DIFFERENT FOR EACH APPLICATION

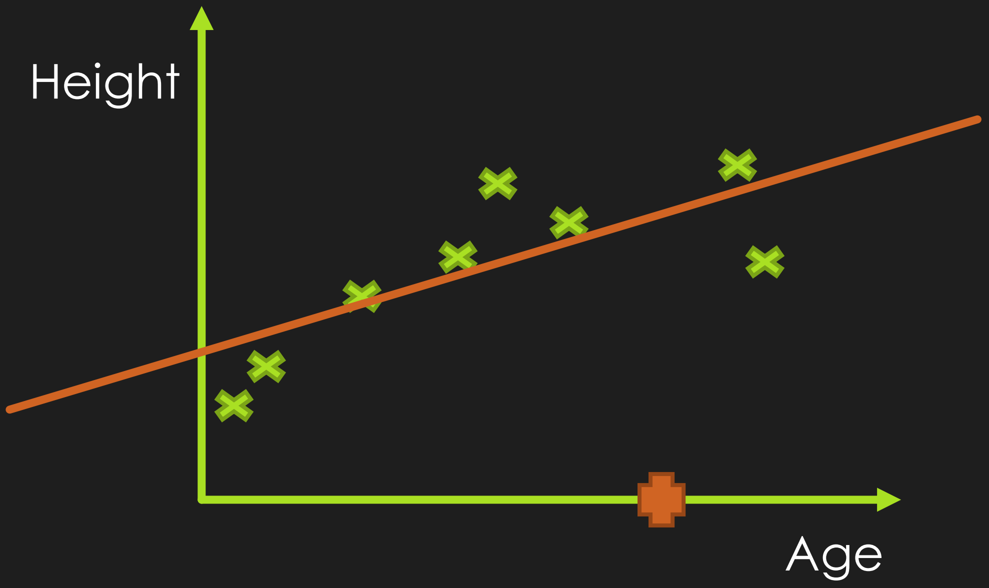






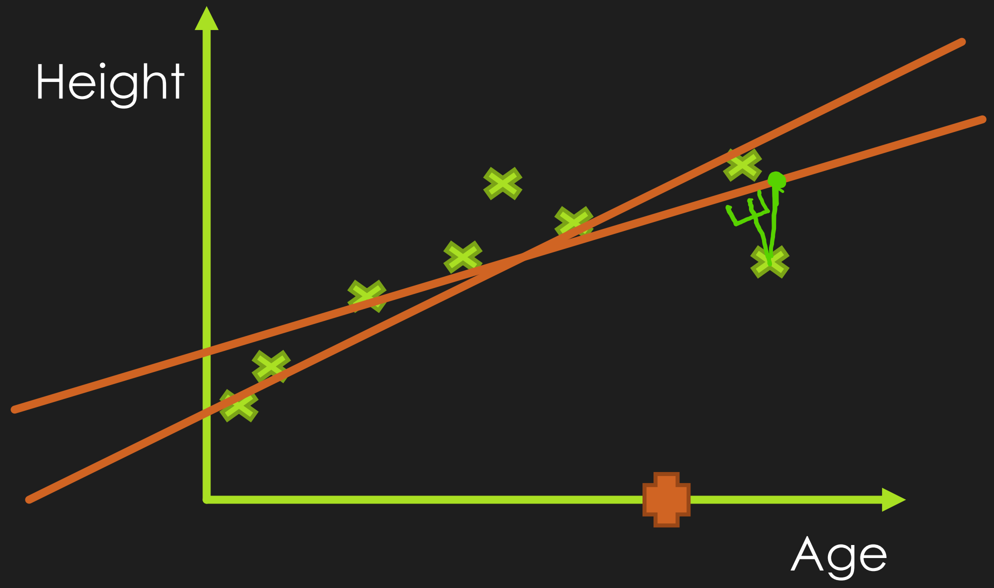
LINEAR REGRESSION

- PREDICTION WHERE
  - $x \in \mathbb{R}^d$
  - $y \in \mathbb{R}$ (COULD BE $\mathbb{R}^k$)
- THE CASE $x, y \in \mathbb{R}$ IS CALLED SIMPLE LINEAR REGRESSION
- BEST WAY OF FITTING A LINE?



LINEAR REGRESSION

- Which line is better?
- Maybe the one that "fits the data" better?

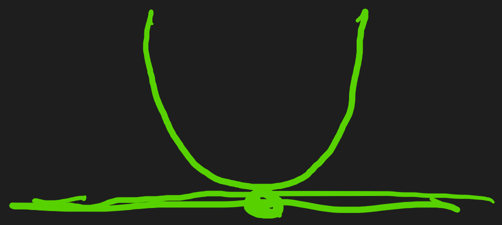


LINEAR REGRESSION

- $\hat{y}^i$ IS THE PREDICTION OF THE MODEL
- LET $d^i = |y^i - \hat{y}^i|$
- BEST LINE MINIMIZES $\sum_{i=1}^{n} (d^i)^2$?
- OTHER OPTIONS?

$$\min \sum_{i=1}^{n} (d^i)^2$$

$y=x^2$



LINEAR REGRESSION

- Ordinary Least Squares (OLS) Method
  - minimize $\sum_{i=1}^{n}(d^i)^2 = \sum_{i=1}^{n}(y^i - \widehat{y^i})^2$
- How about?
  - Minimize $\sum_{i=1}^{n} |y^i - \widehat{y^i}|$
  - Minimize $\sum_{i=1}^{n}\left(\frac{y^i}{\widehat{y^i}} + \frac{\widehat{y^i}}{y^i}\right)$ or Minimize $\sum_{i=1}^{n}\left(\frac{y^i+a}{\widehat{y^i}+a} + \frac{\widehat{y^i}+a}{y^i+a}\right)$
  - Minimize $\sum_{i=1}^{n} \log\left|\frac{y^i+0.0001}{\widehat{y^i}+0.0001}\right|$

    Examples: Revenue, etc.
  - …

1-D ORDINARY LEAST SQUARES

- $x, y \in \mathbb{R}$
- FIND $a, b \in \mathbb{R}$ SUCH THAT $\hat{y} = ax + b \approx y$
- WE ARE GIVEN $\{(x^i, y^i)\}_{i=1}^{n}$
- FIND/LEARN a, b FROM THE DATA

$$\min_{a,b} \sum_{i=1}^{n} \left(\underbrace{ax^i + b}_{\hat{y}^i} - y^i\right)^2$$

- MINIMIZE $f(a,b) = \sum_{i=1}^{n} \left(ax^i + b - y^i\right)^2$

$$\frac{\partial f}{\partial a} = 2 \sum_{i=1}^{n} \left(ax^i + b - y^i\right) x^i = 0$$

$$\frac{\partial f}{\partial b} = 2 \sum_{i=1}^{n} \left(ax^i + b - y^i\right) = 0$$

$$\Rightarrow 2nb + 2a \sum_{i=1}^{n} x^i = 2 \sum_{i=1}^{n} y^i$$

$$\Rightarrow b + a\bar{x} = \bar{y} \Rightarrow \boxed{b = \bar{y} - a\bar{x}}$$

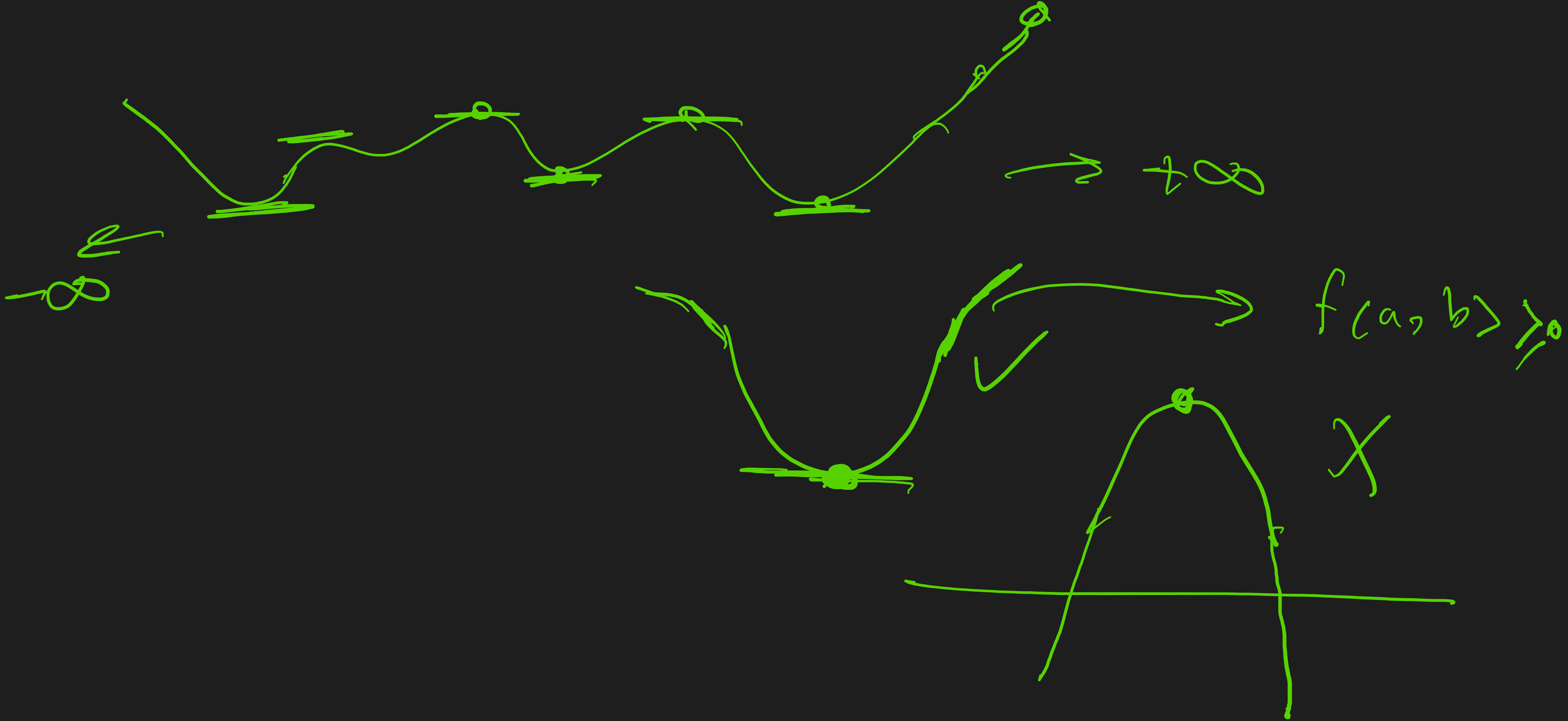
$$\bar{y} = \frac{1}{n} \sum y^i$$

$$\bar{x} = \frac{1}{n} \sum x^i$$

1-D ORDINARY LEAST SQUARES

- OPTIMAL a AND b:

- $\boxed{a = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2}} = \frac{COV(x,y)}{Var(x)}, \underline{b = \bar{y} - a\bar{x}}$
  - $\bar{x} = \frac{1}{N}\sum x^i$, $\bar{y} = \frac{1}{N}\sum y^i$, $\underline{\overline{xy} = \frac{1}{N}\sum x^i y^i}$, $\overline{x^2} = \frac{1}{N}\sum \left(x^i\right)^2$


$-\infty$
$+\infty$
$f(a,b) \geq 0$
✓
X