Advanced Machine Learning For Design

Lecture 4 - Machine Learning for Image Processing (part 1)

Module 2

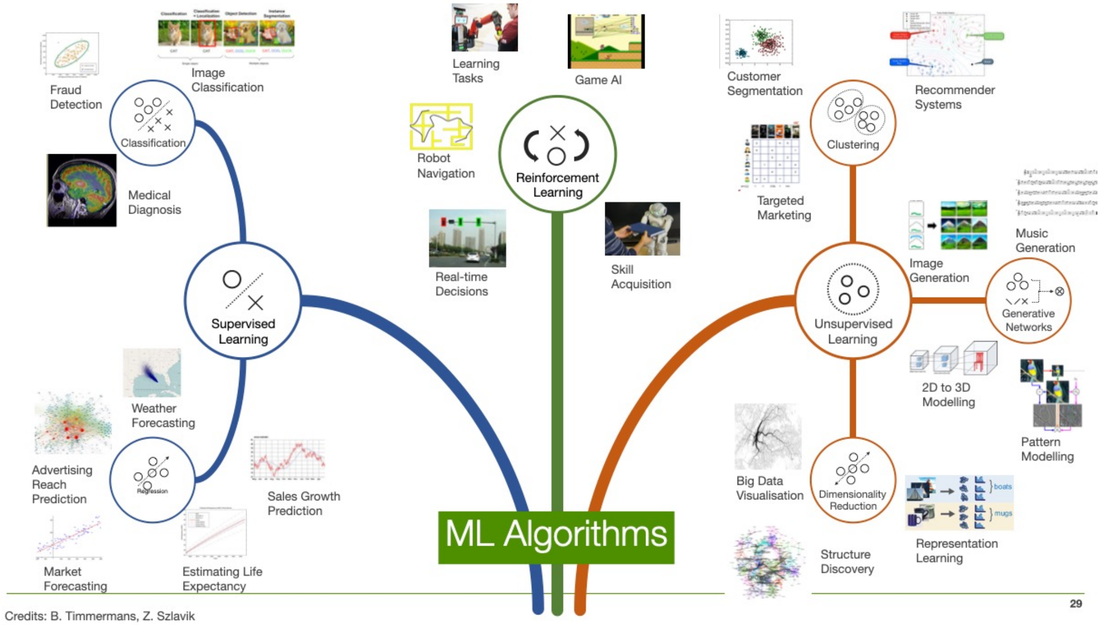
Evangelos Niforatos

12/10/2022

aml4d-ide@tudelft.nl
https://aml4design.github.io/

Admin

Recap on ML/AI….



Credits: B. Timmermans, Z. Szlavik

29

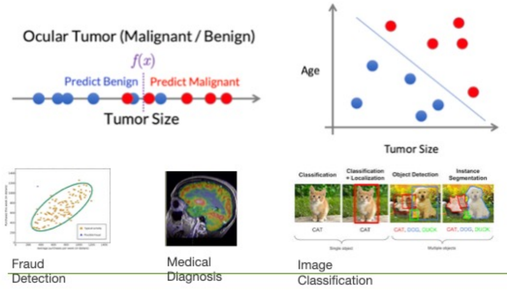
Classification / Regression

Classification

- Learn to output a **category** label
  - Binary (e.g. Spam / not Spam, Cat / not cat)
  - Multi-class (e.g. cat, dog, bird)

Regression (prediction)

- Learn to guess one or more numbers
  - e.g. value of a share, number of stars in a review



33

Images

- Visual content acquired through cameras, scanners, etc.
- Each pixel in an image is a feature
  - But spatially and geometrically organised
  - e.g. edges, corners
- Feature values are numerical values across channels
  - e.g. R,G,B
- Dimensionality —> n x m

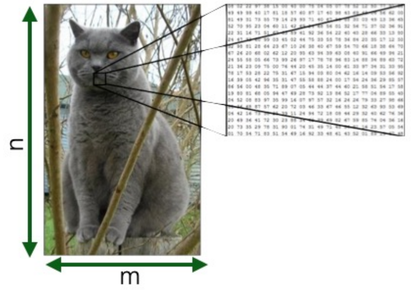


Image →

P(1,1)	P(2,1)	P(3,1)	...	P(n,m)	Class
255, 0, 0	255, 1, 1	255, 0, 0		R,G,B	Cat
255, 213, 0	255, 213, 1	255, 213, 4		R,G,B	Dog
					Cat
					Duck

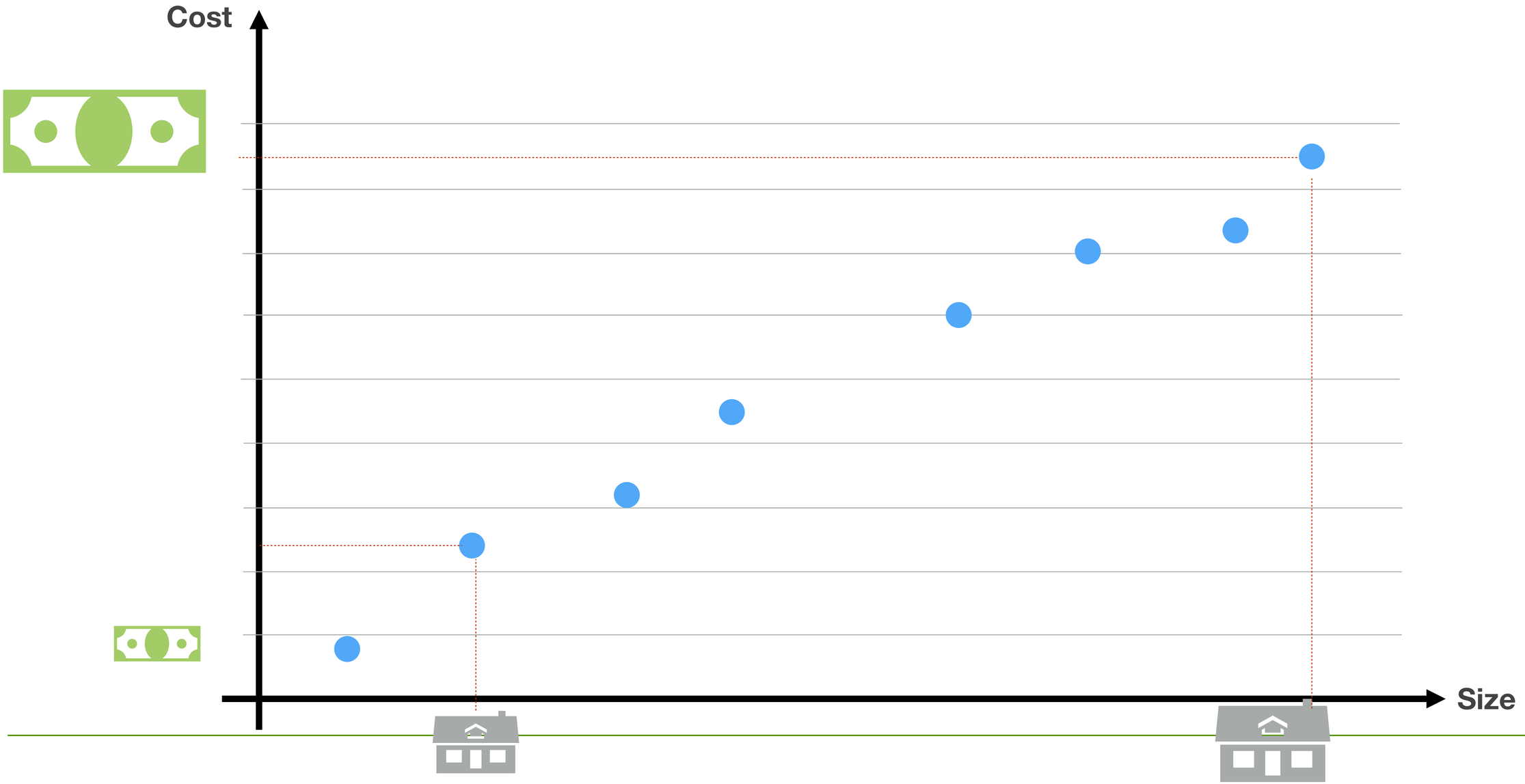
More in Module 2

21

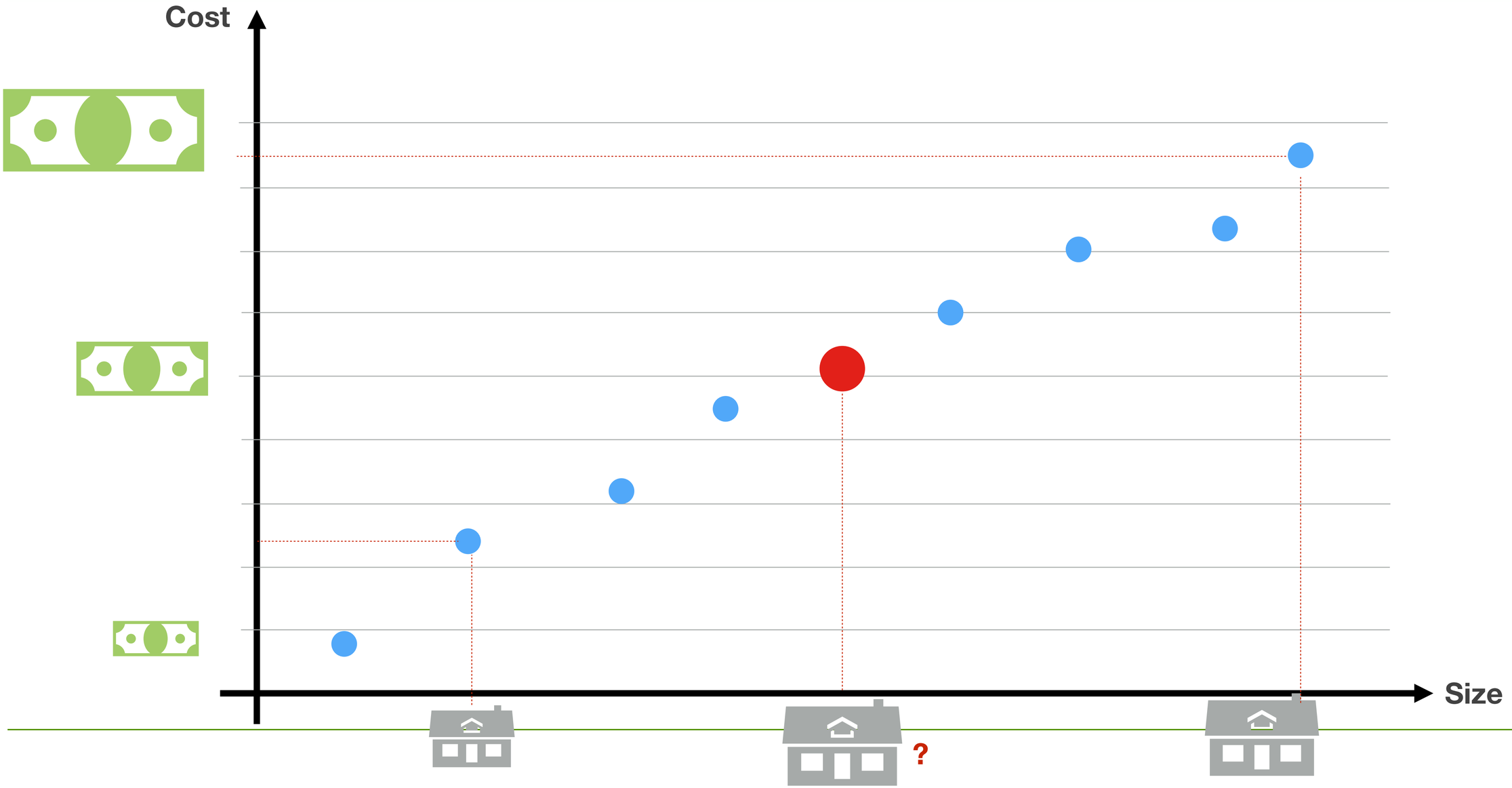
A bit more on regression and classification

And your 2nd contact with (deep) neural networks

Linear Regression /1



Linear Regression /2

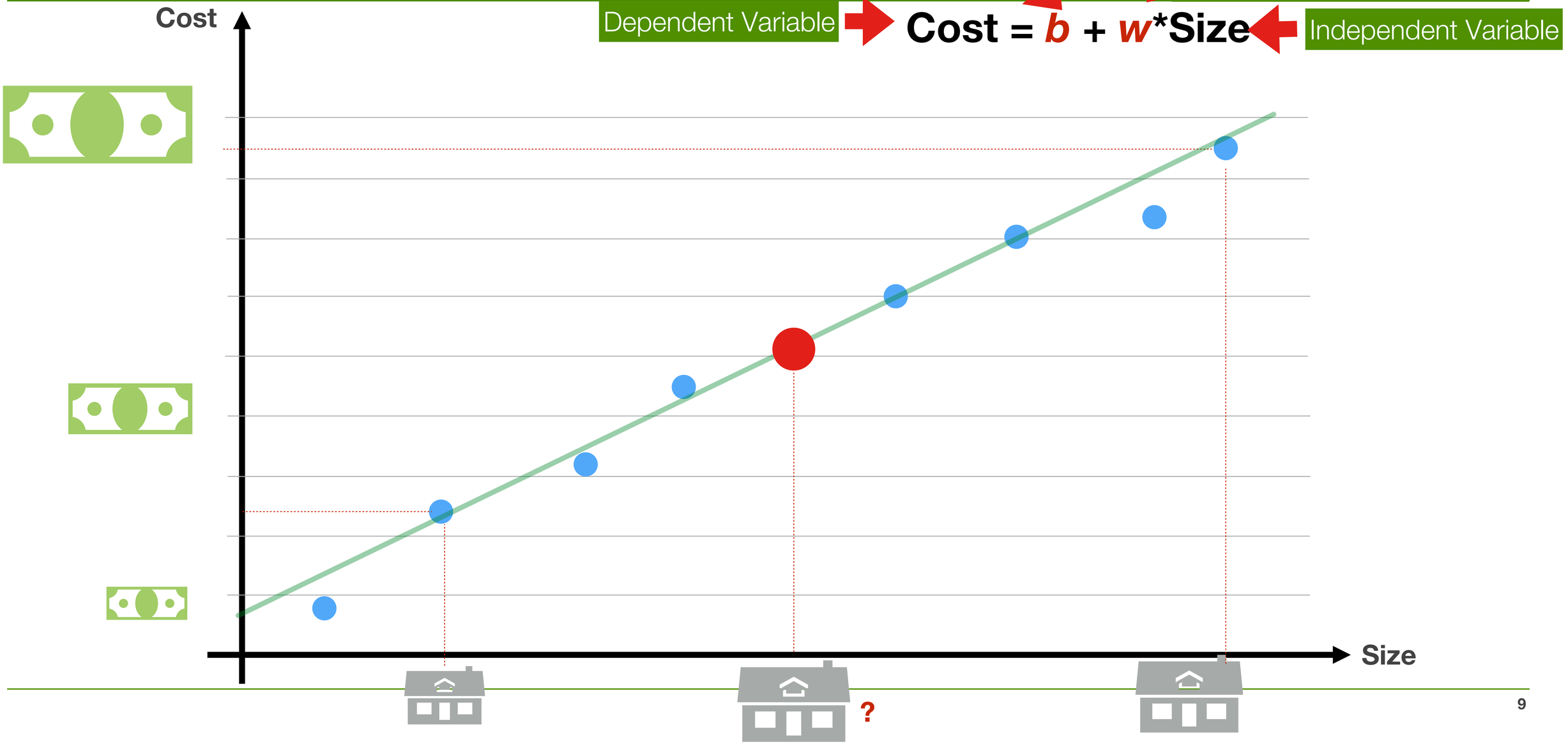


Linear Regression

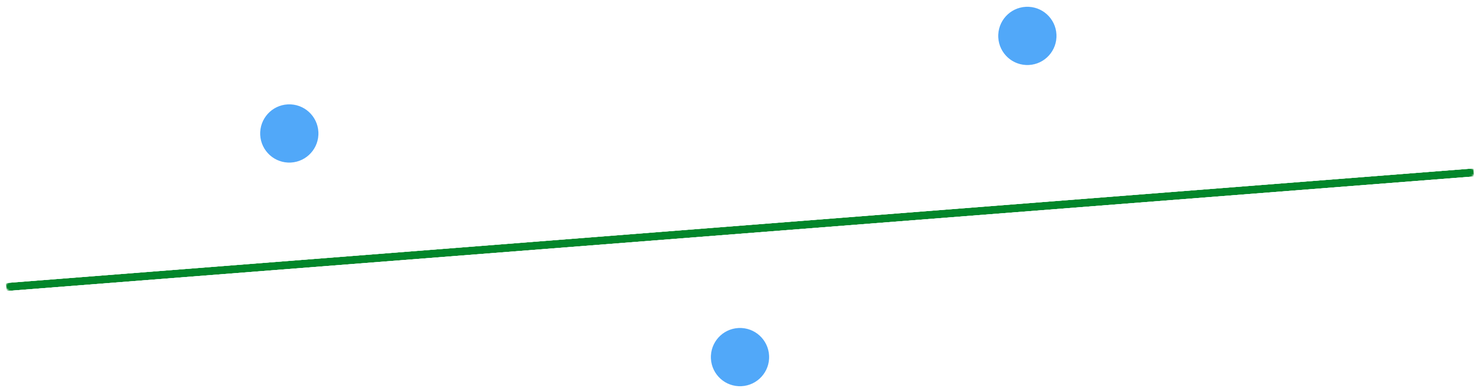
Intercept (parameter, or **bias**)

Slope (parameter, or **weight**)

Dependent Variable → **Cost = *b* + *w**Size** ← Independent Variable



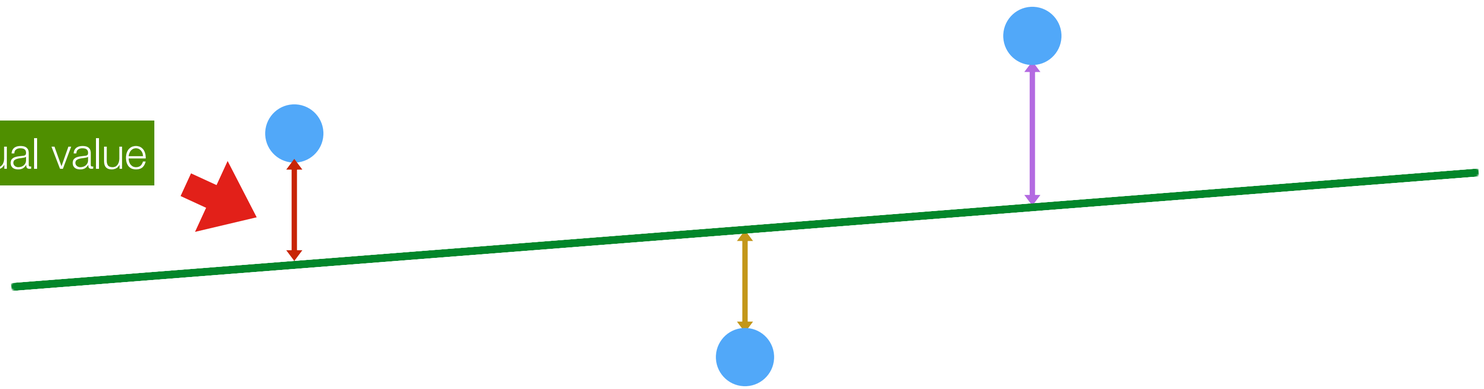
Cost = *w**Size



Cost = *wSize**

predicted value - actual value

Fit$_1$:

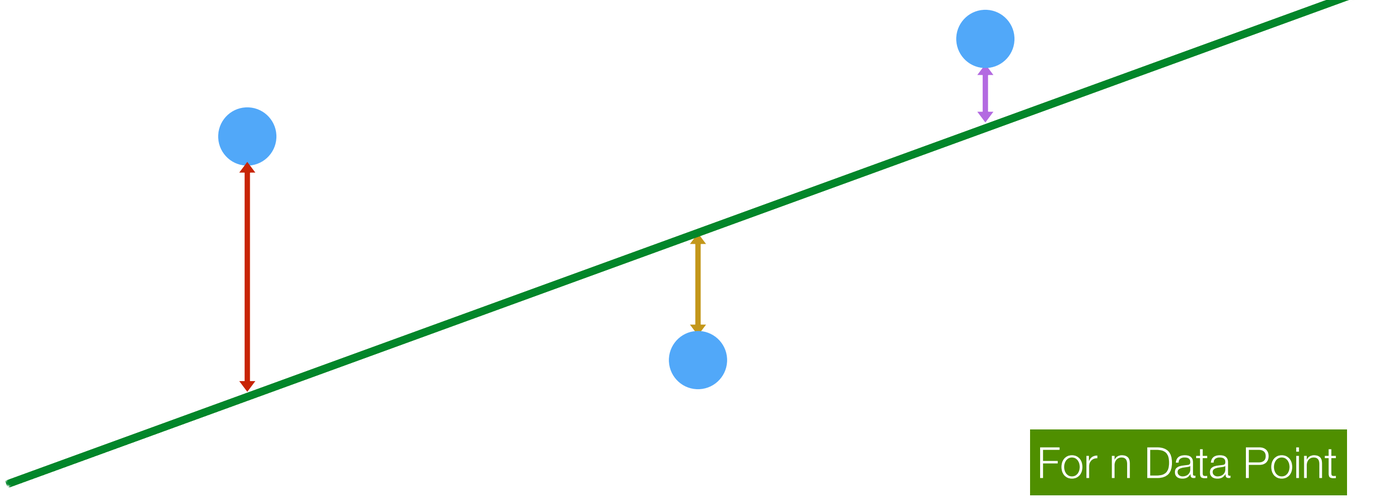


***w* = 0.05**

Error$_1$:

Cost = *wSize**

***w* = 0.2**



Fit$_2$:

For n Data Point

Error$_1$:

Error$_2$:

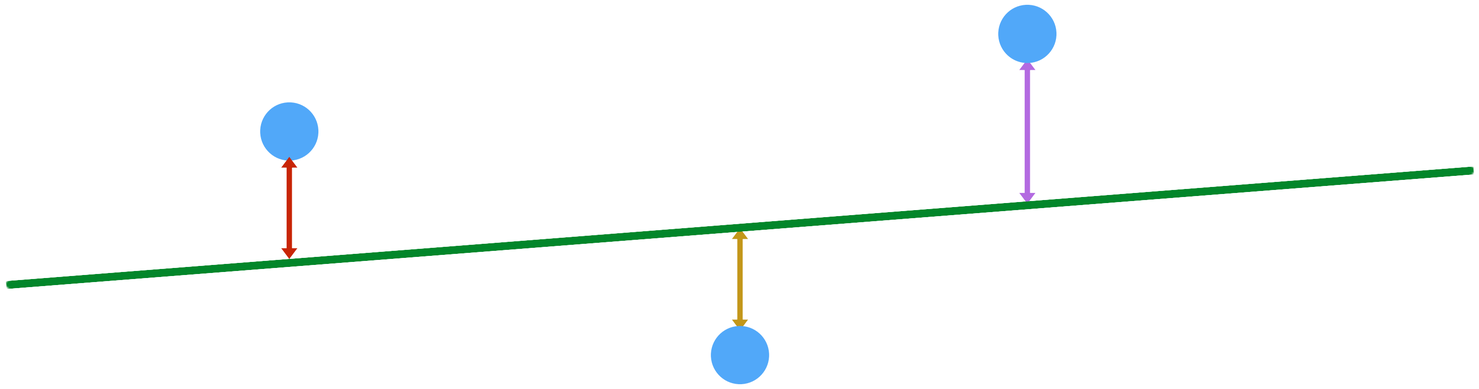
$$Error = \frac{1}{n}\sum_{i=1}^{n}(Prediction_i - Value_i)^2$$

Mean Squared Error (MSE)

Current data point

Cost = *wSize**

Fit$_1$:



***w* = 0.05**

Error$_1$:

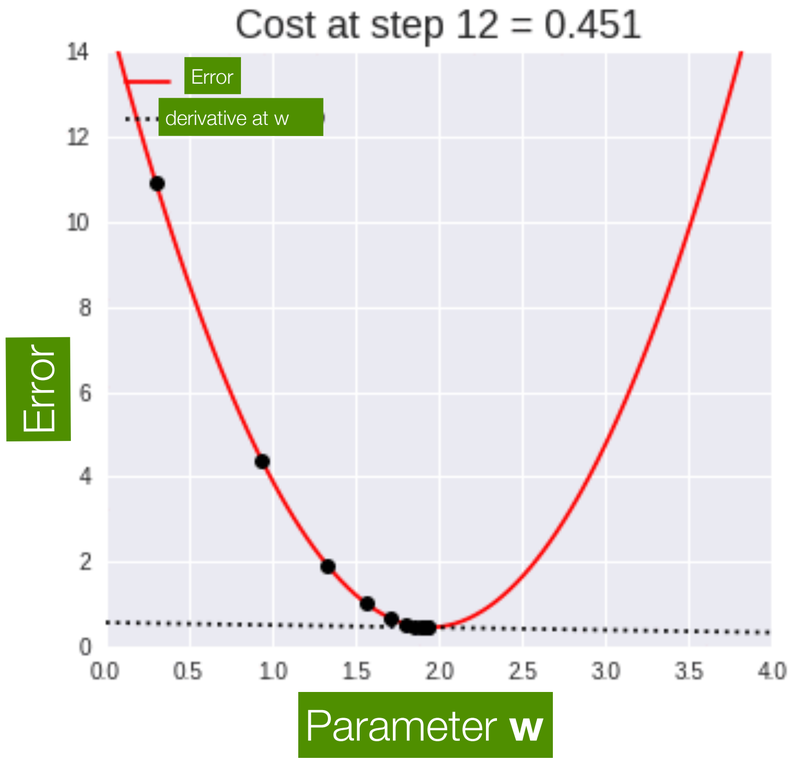
Error$_2$:

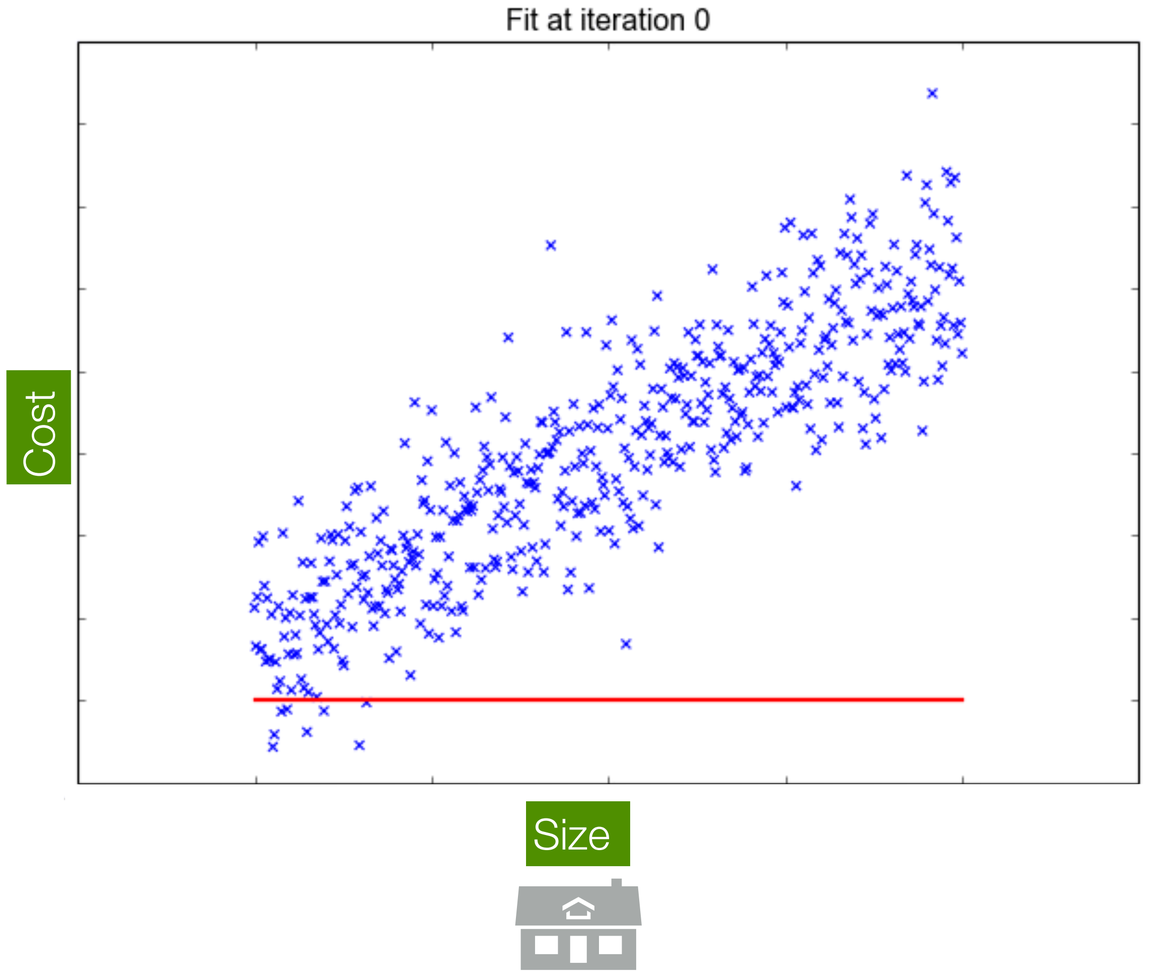
Fit$_1$ is a better fit on the training data than Fit$_2$

We select ***w* = 0.05**

Finding the best parameter values (training the model)

- **Gradient descent**

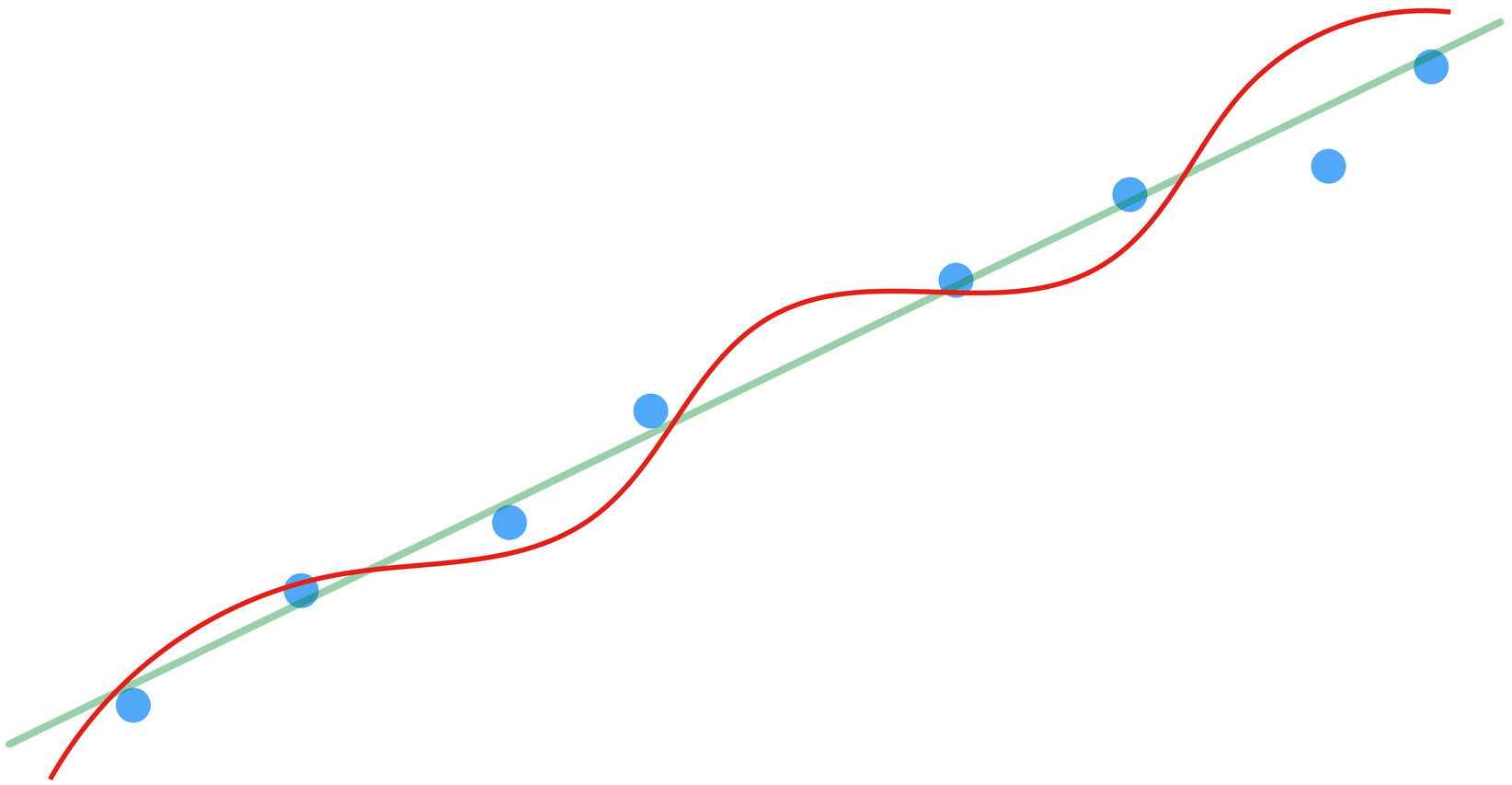




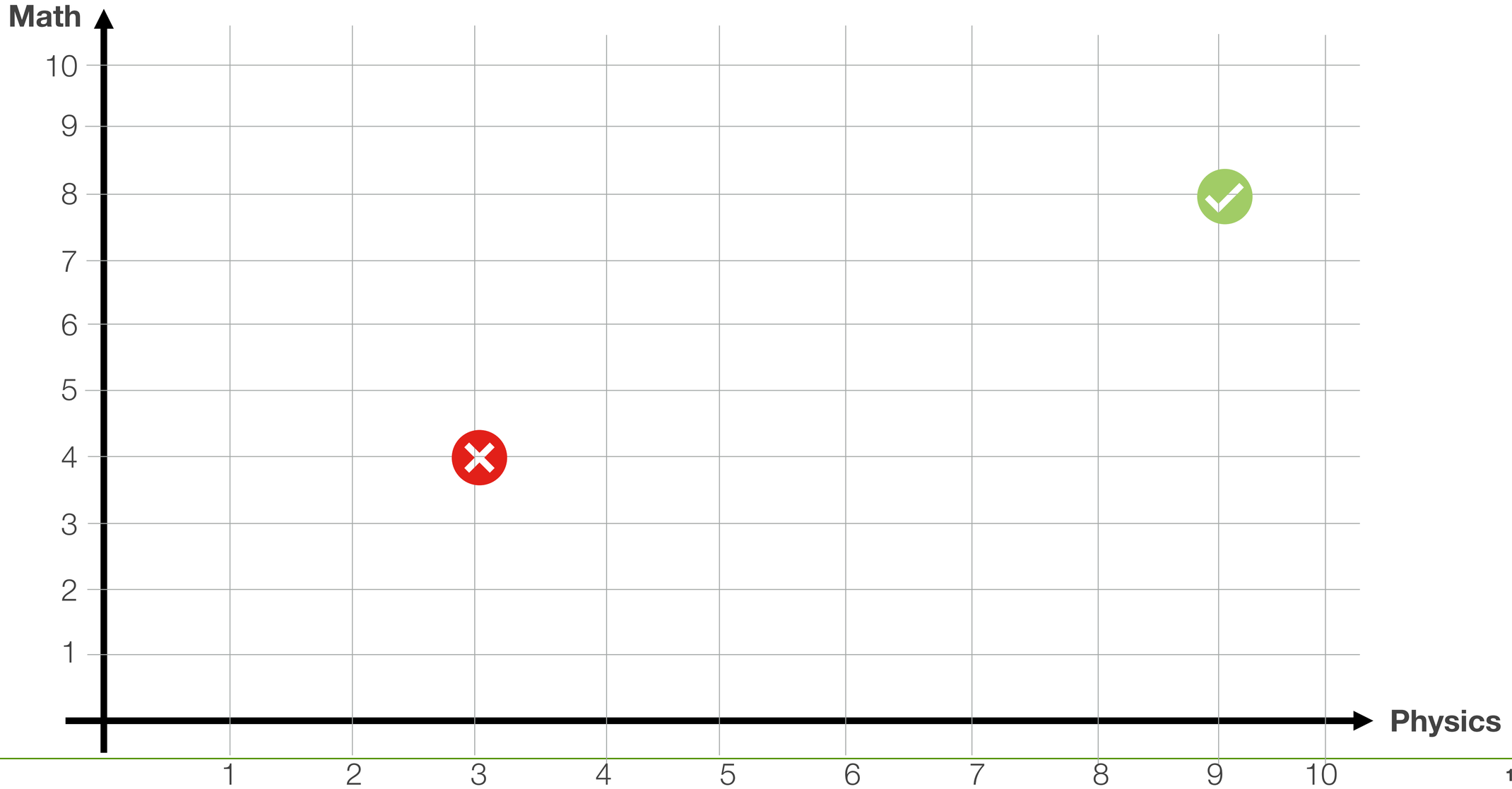
- **Hyperparameters**
  - **Learning Rate**: “speed” of descent
  - **Epochs**: max number of steps

Polynomial Regression

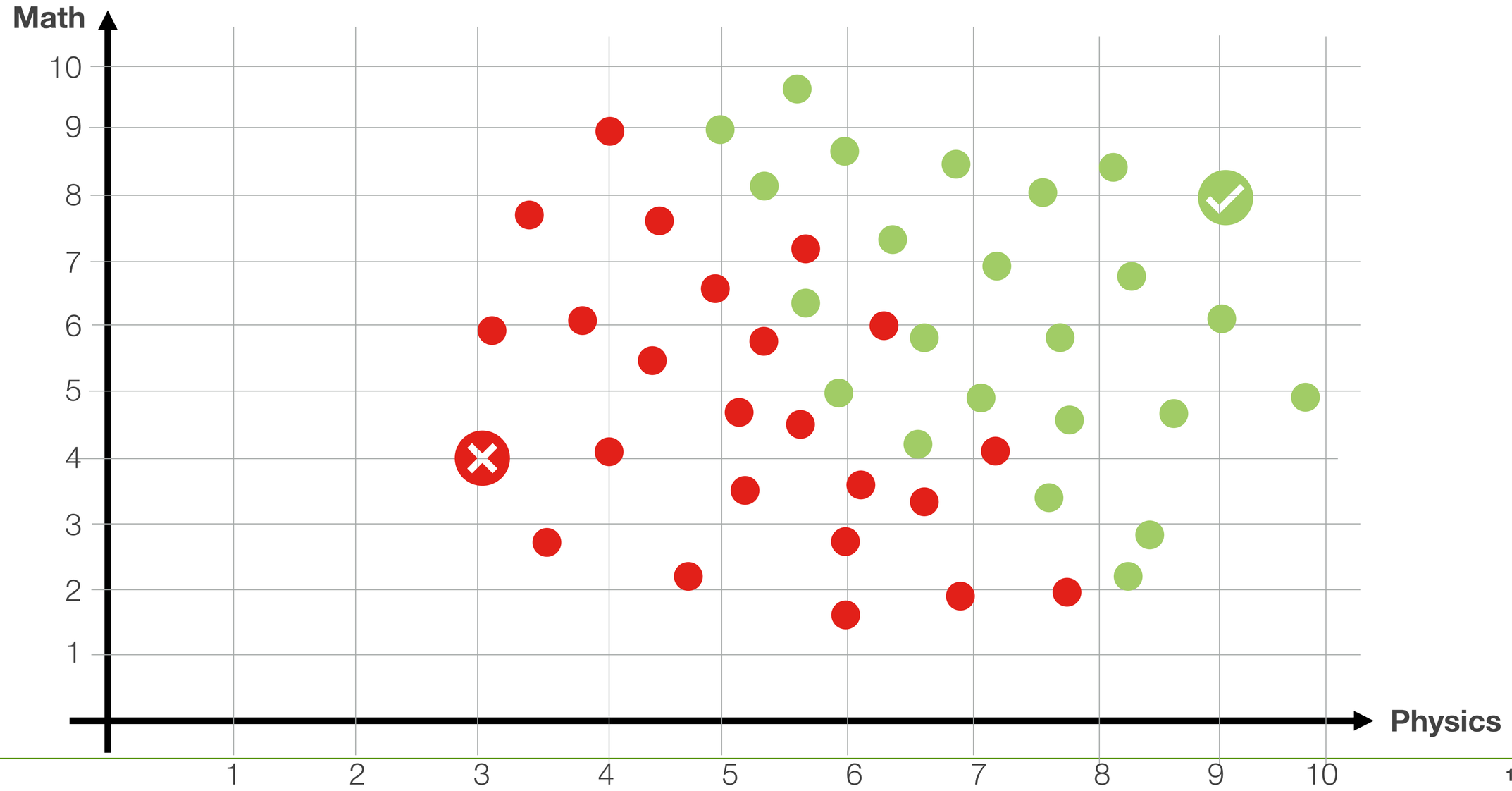
N^{th} degree polynomial → **Cost = b + w_1 Size + w_2 Size^2 + …. + w_n Size^n**



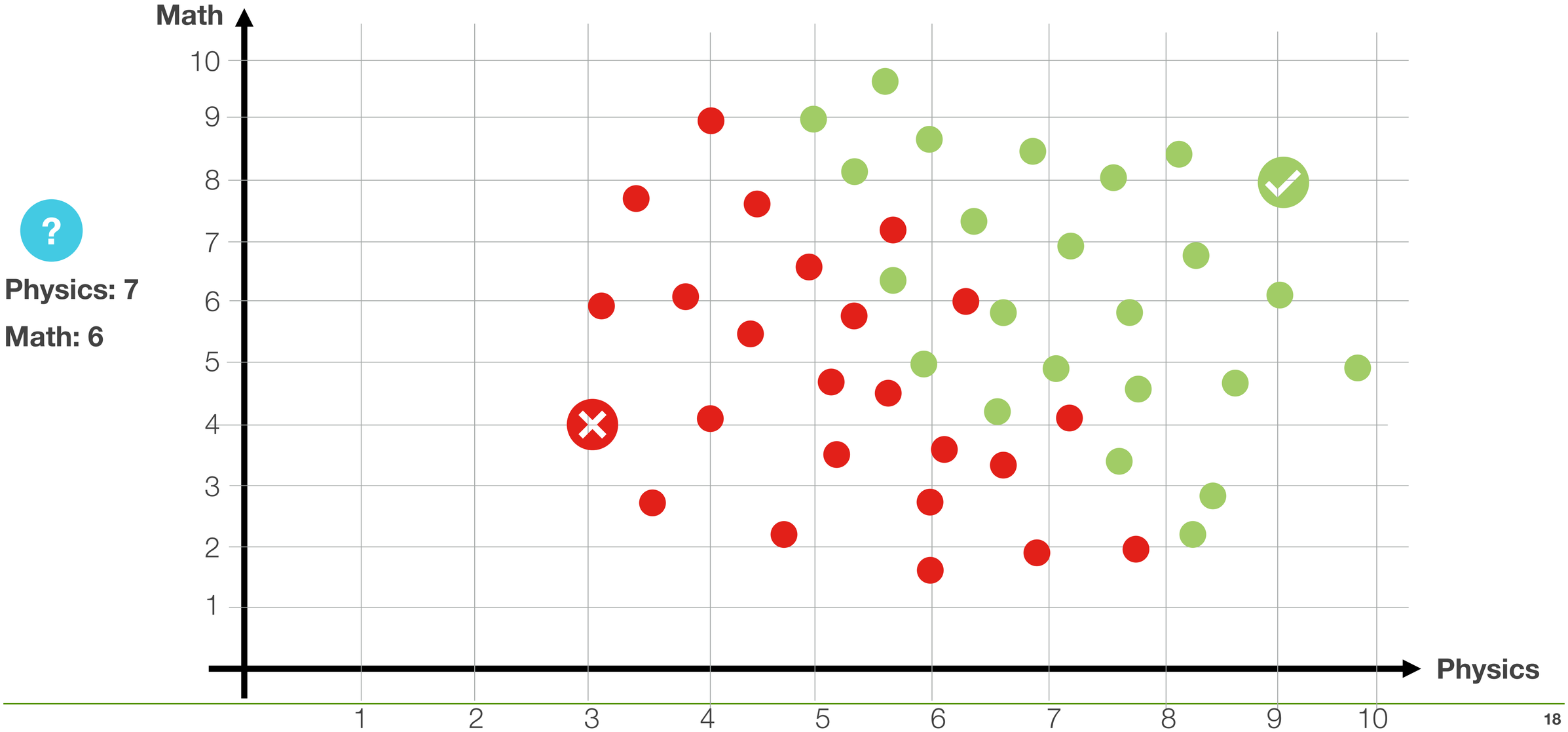
Classification



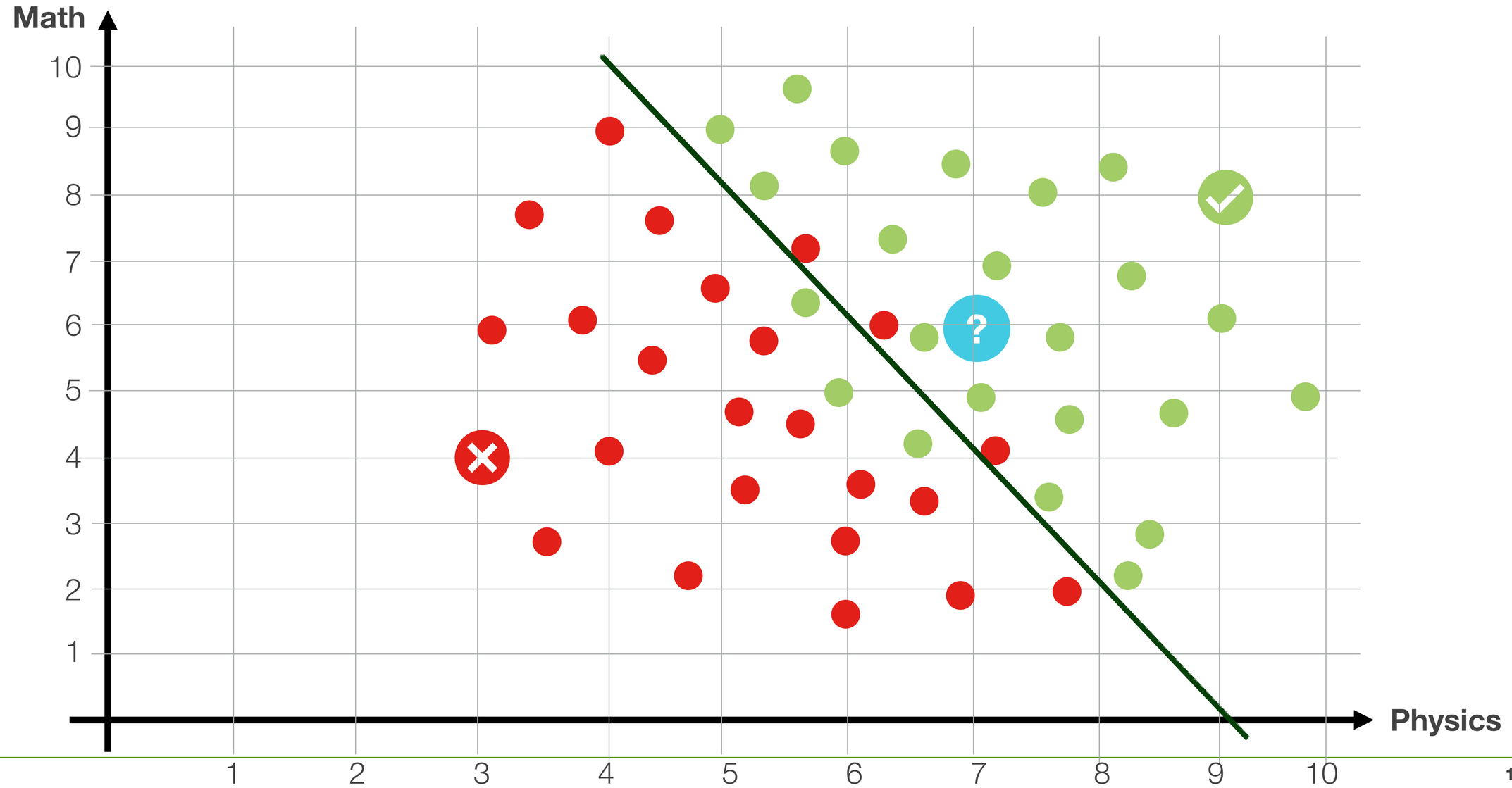
Classification

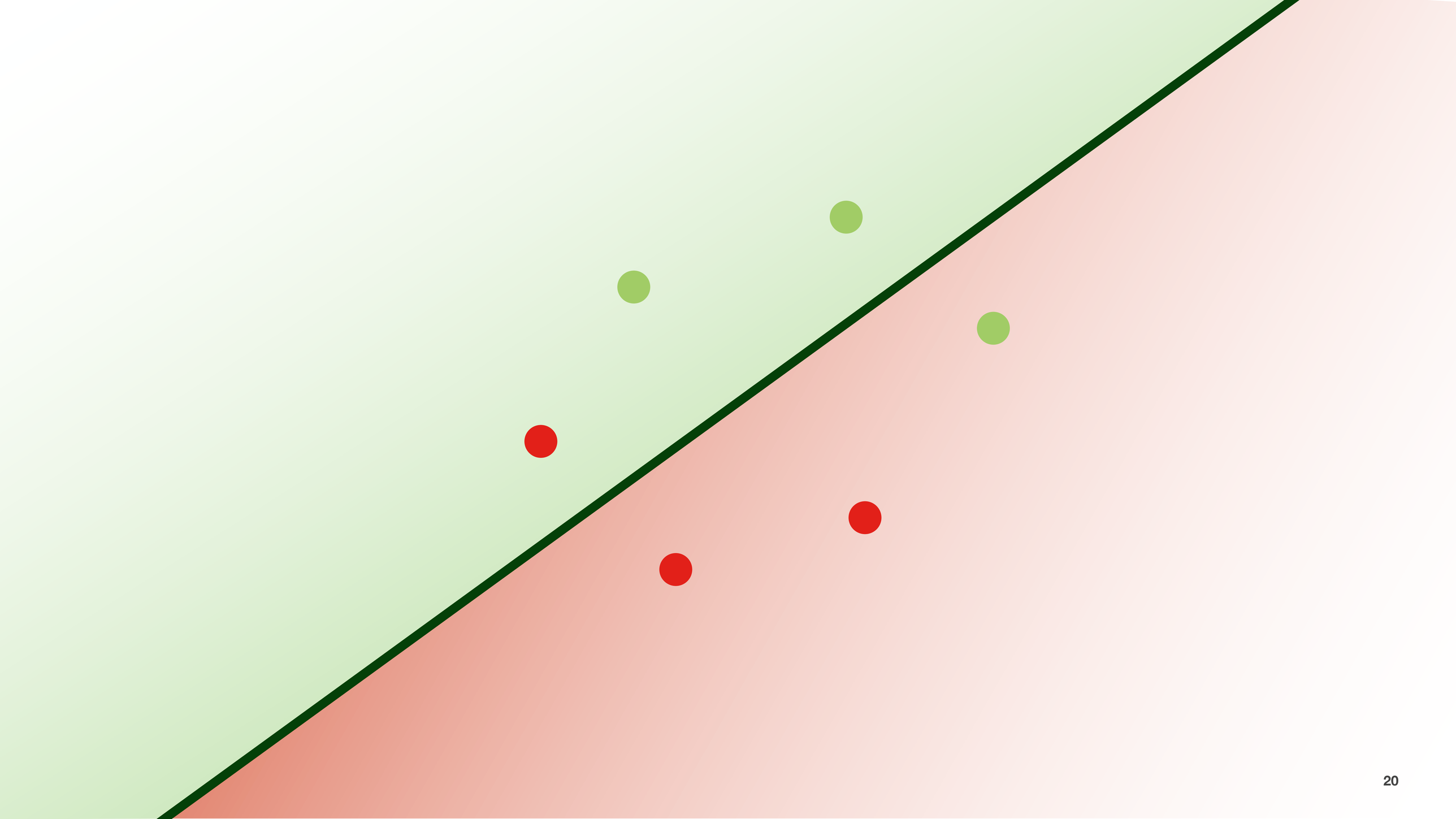


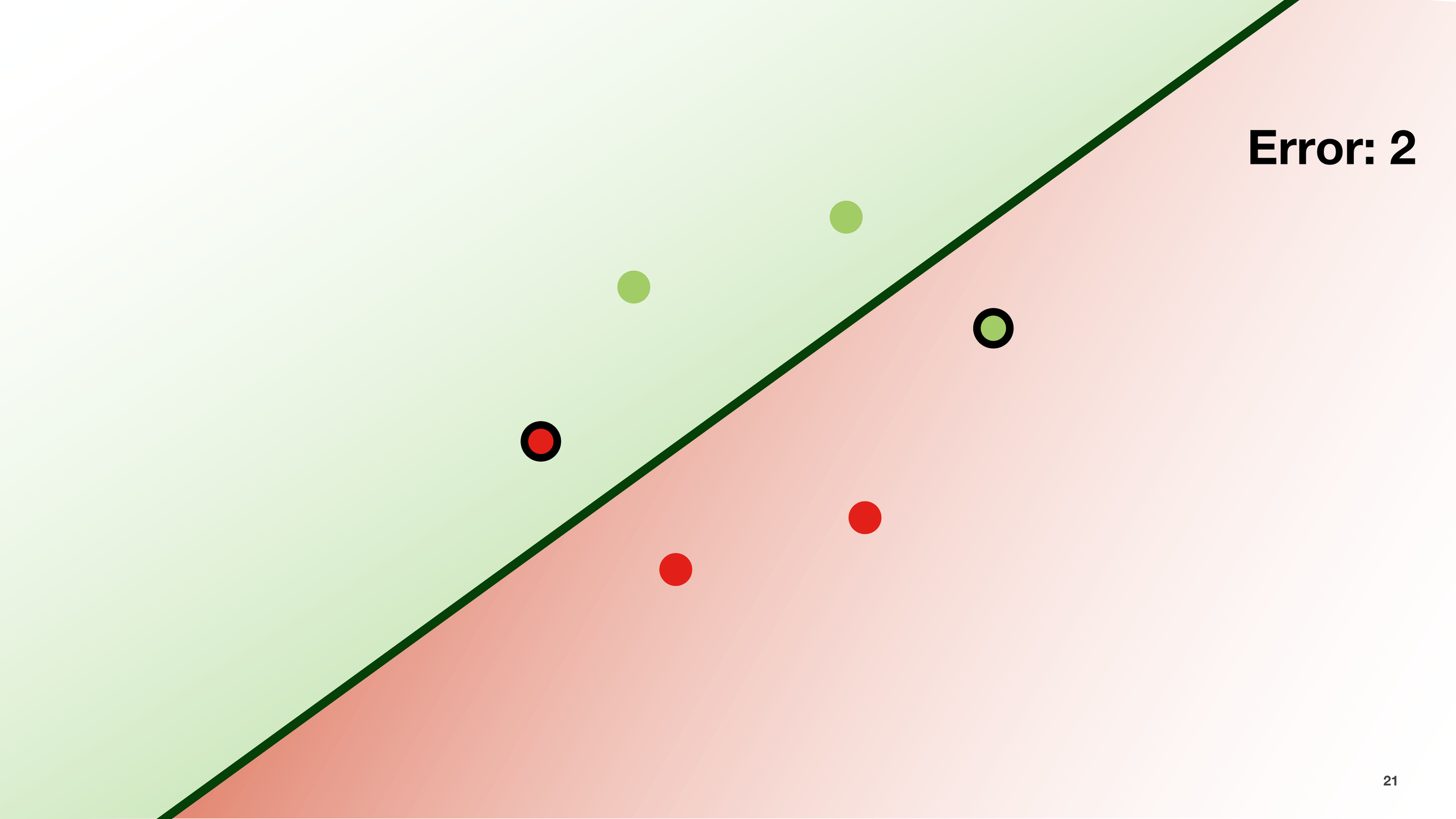
Classification

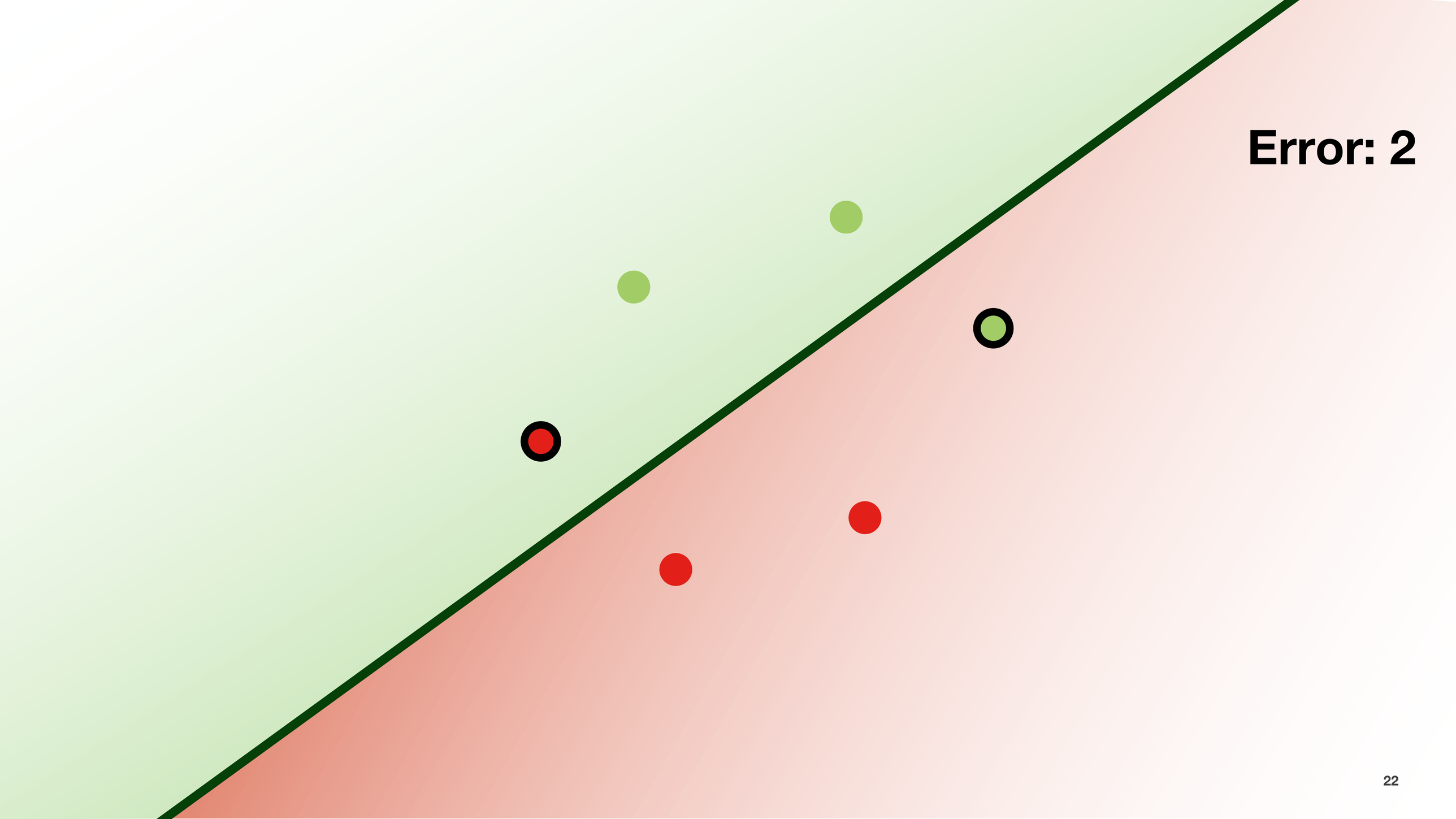


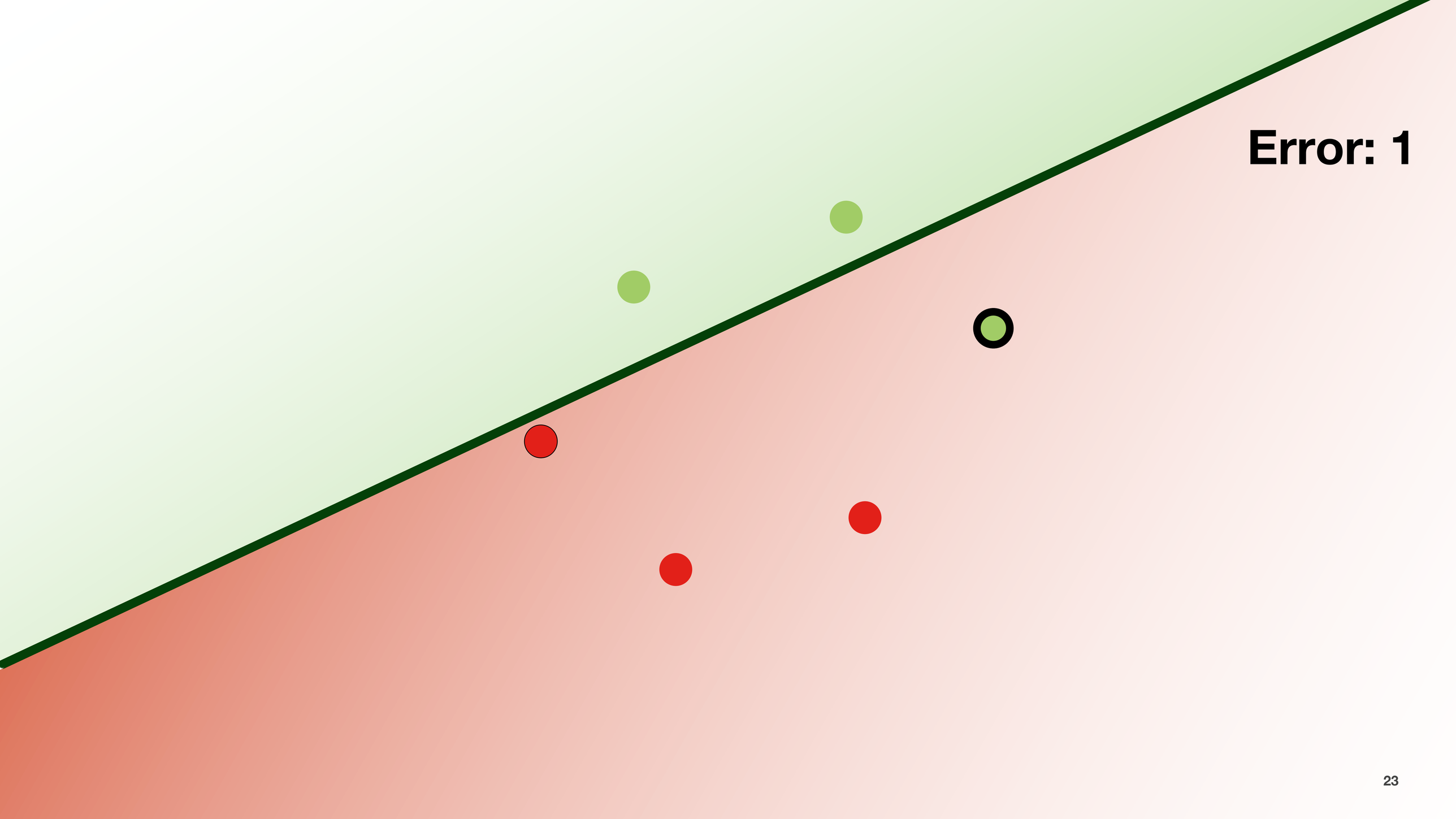
Logistic Regression

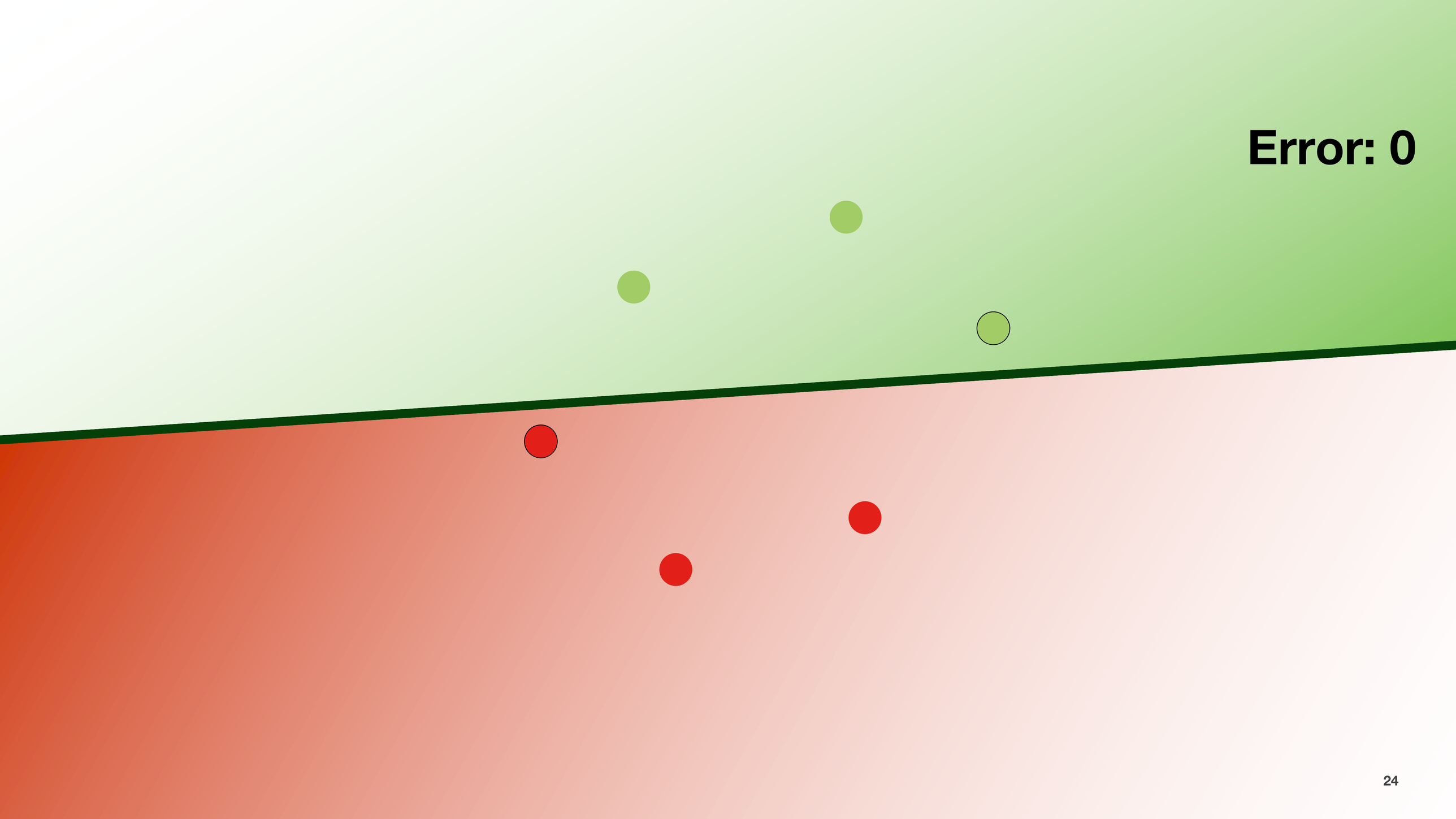


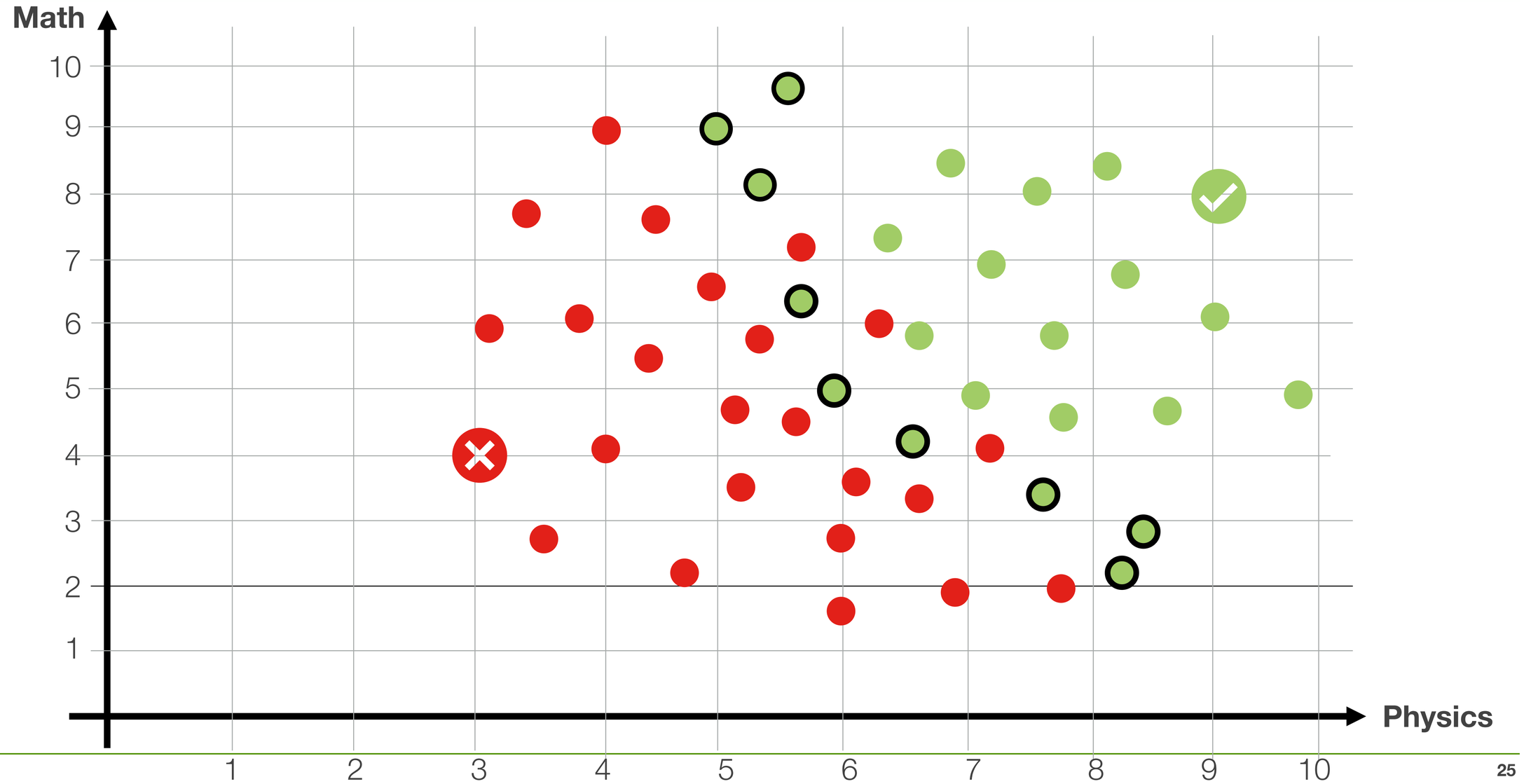


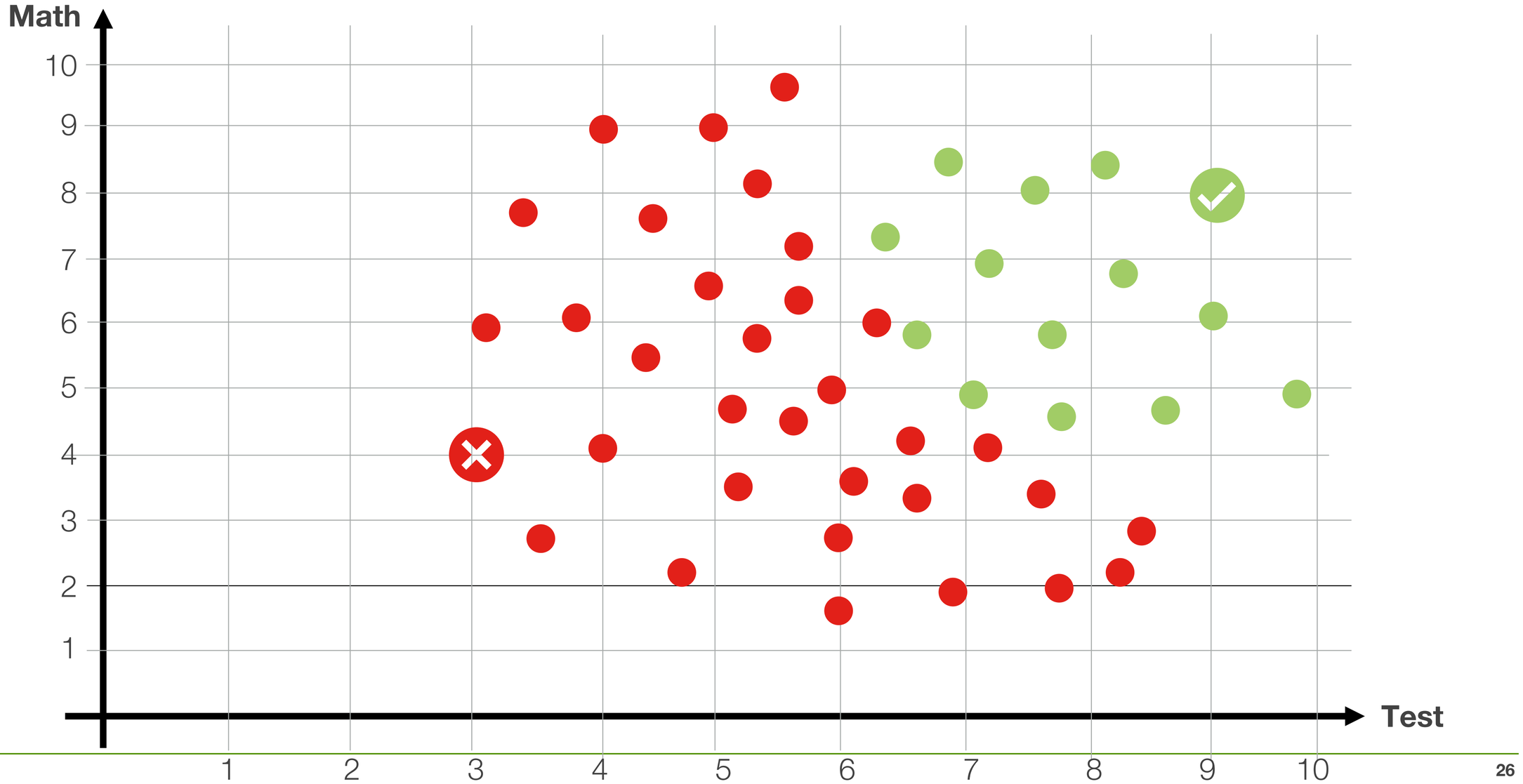

Error: 2

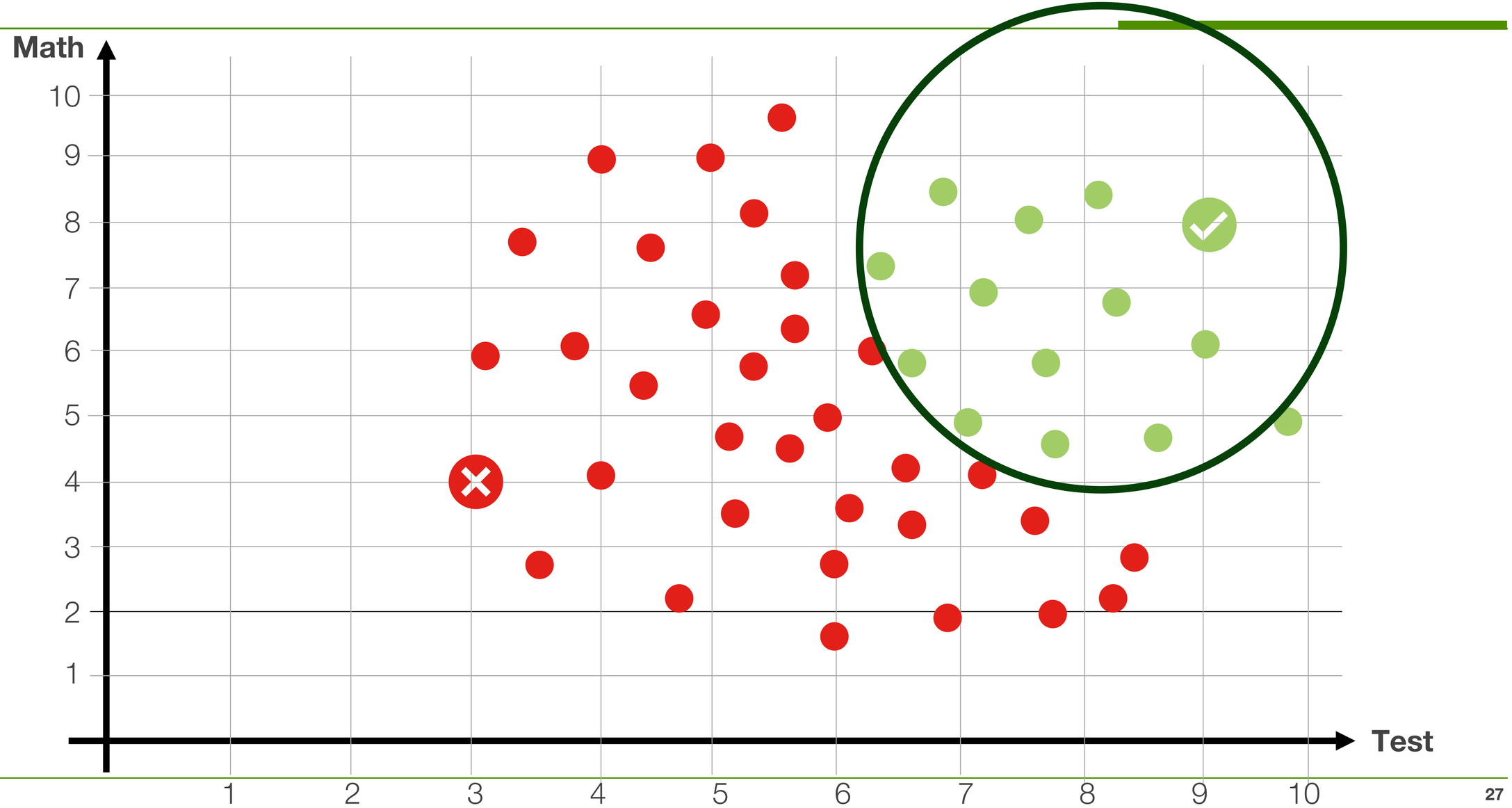

Error: 2

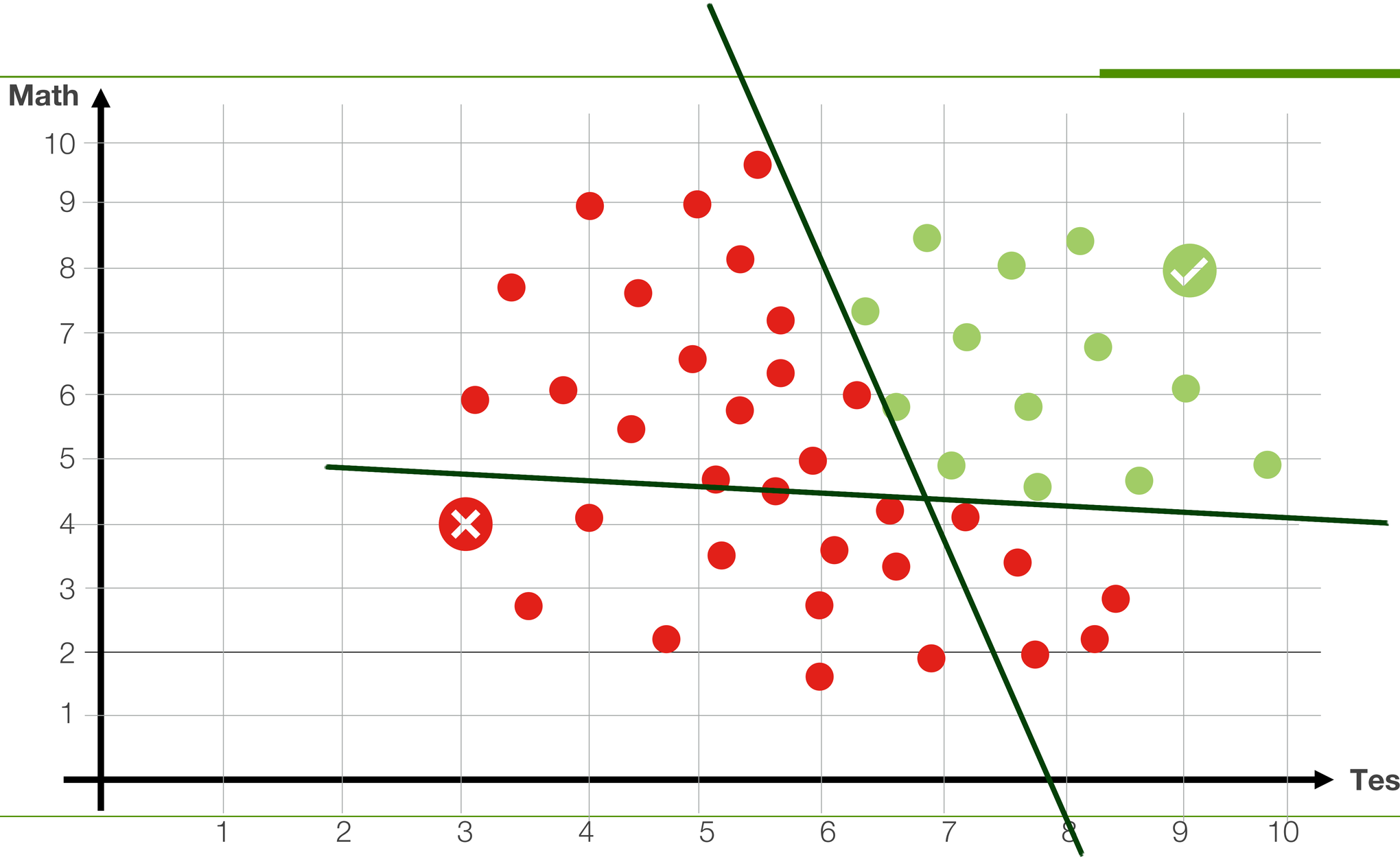

Error: 1


Error: 0

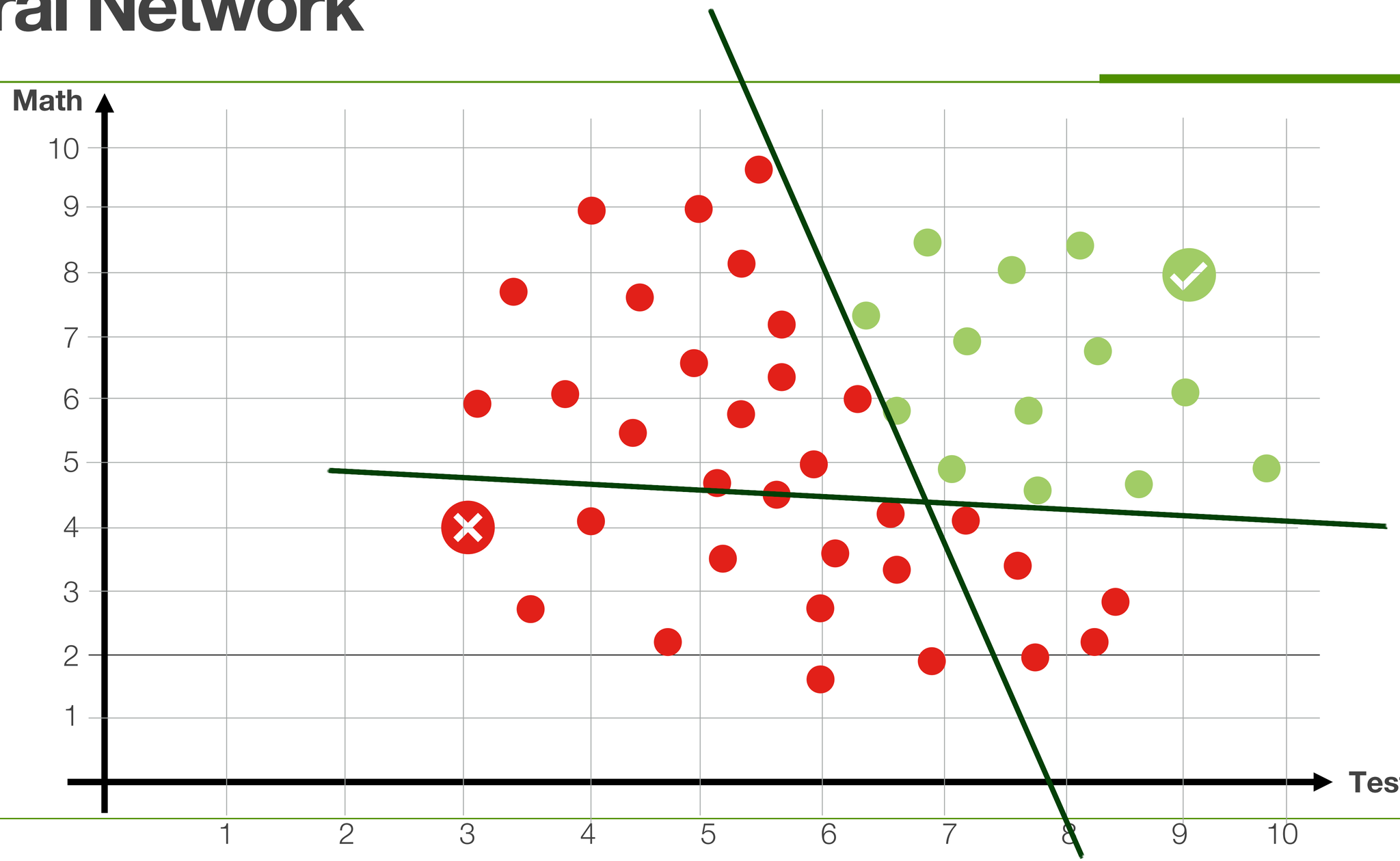

Math
10
9
8
7
6
5
4
3
2
1
Physics
1
2
3
4
5
6
7
8
9
10

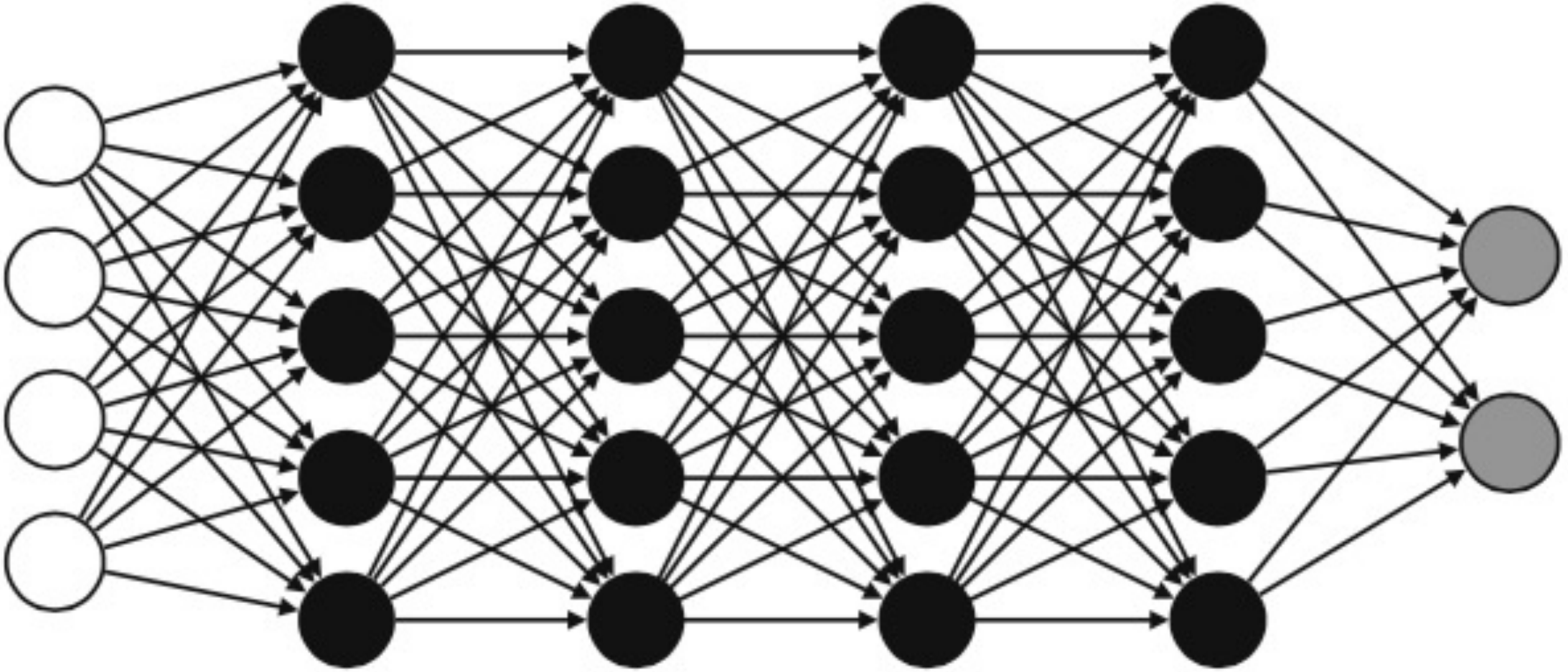

Math
10
9
8
7
6
5
4
3
2
1
Test
1
2
3
4
5
6
7
8
9
10


Math
10
9
8
7
6
5
4
3
2
1
Test
1
2
3
4
5
6
7
8
9
10

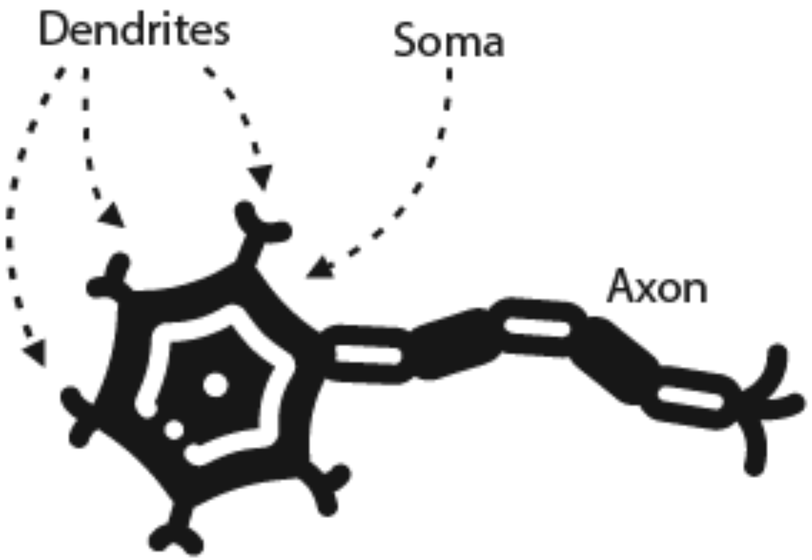

Math
10
9
8
7
6
5
4
3
2
1
Test
1
2
3
4
5
6
7
8
9
10

Neural Network

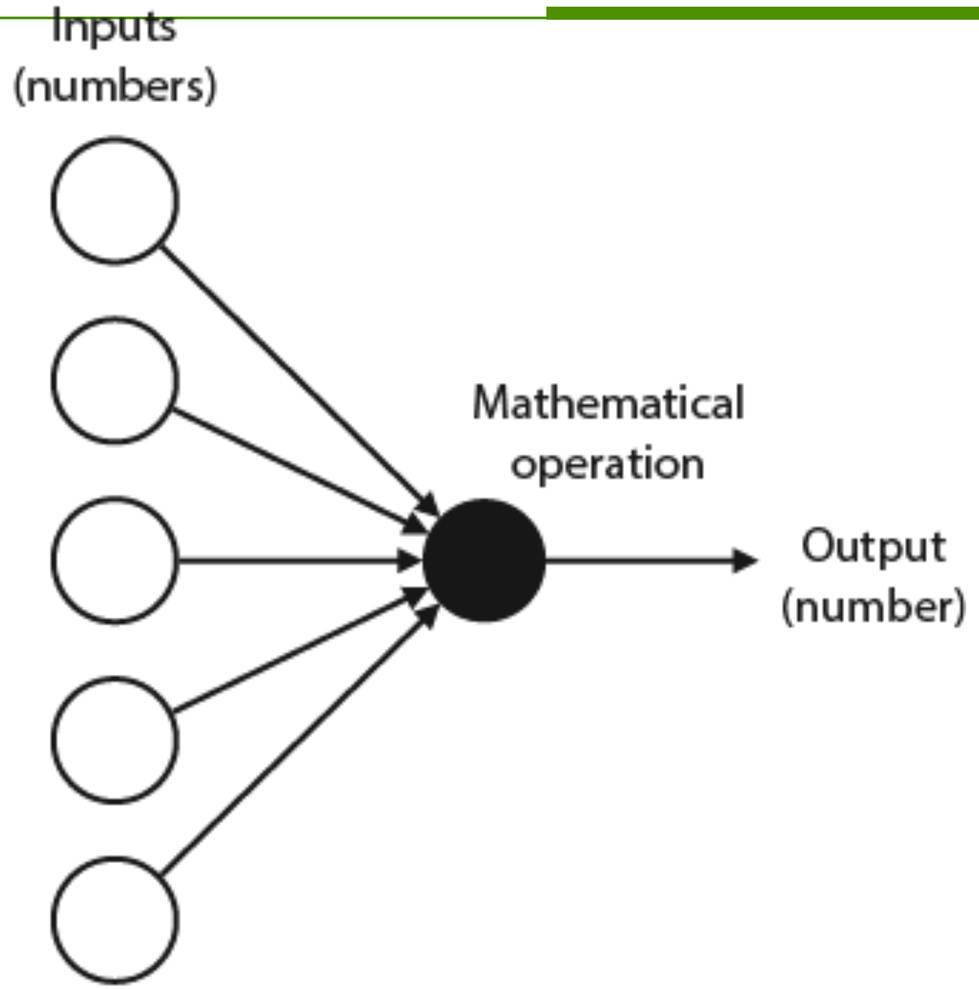




Neural Network

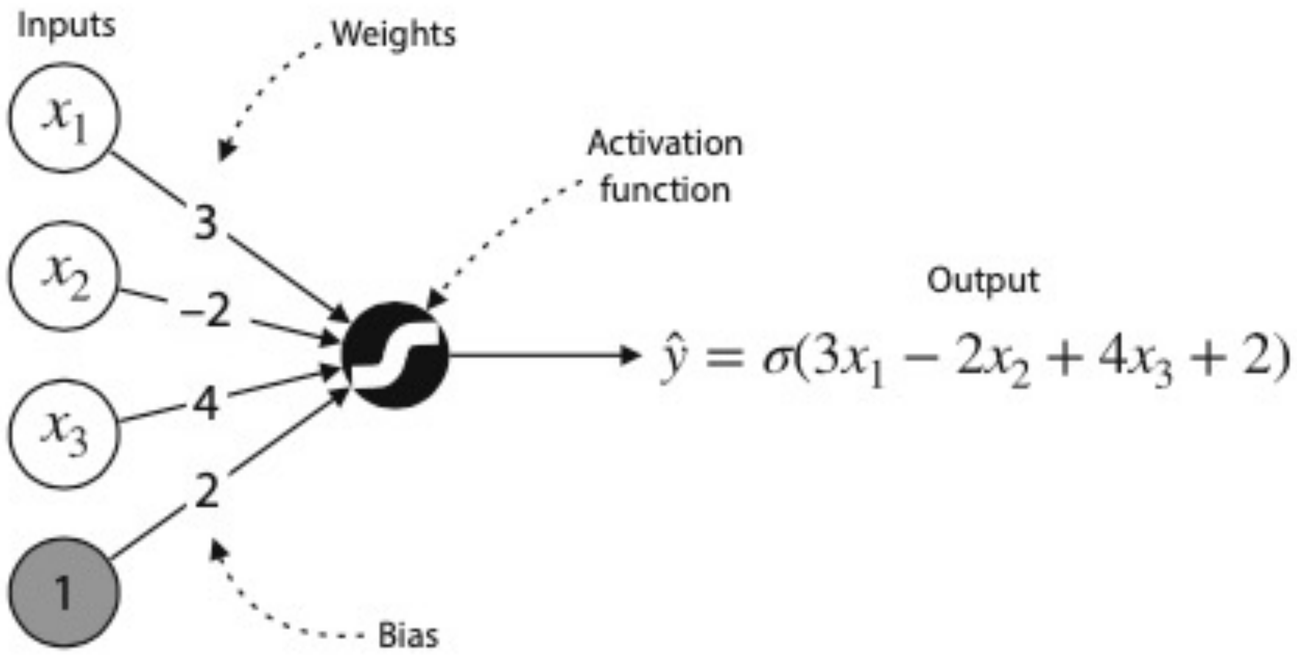


Neuron



Perceptron

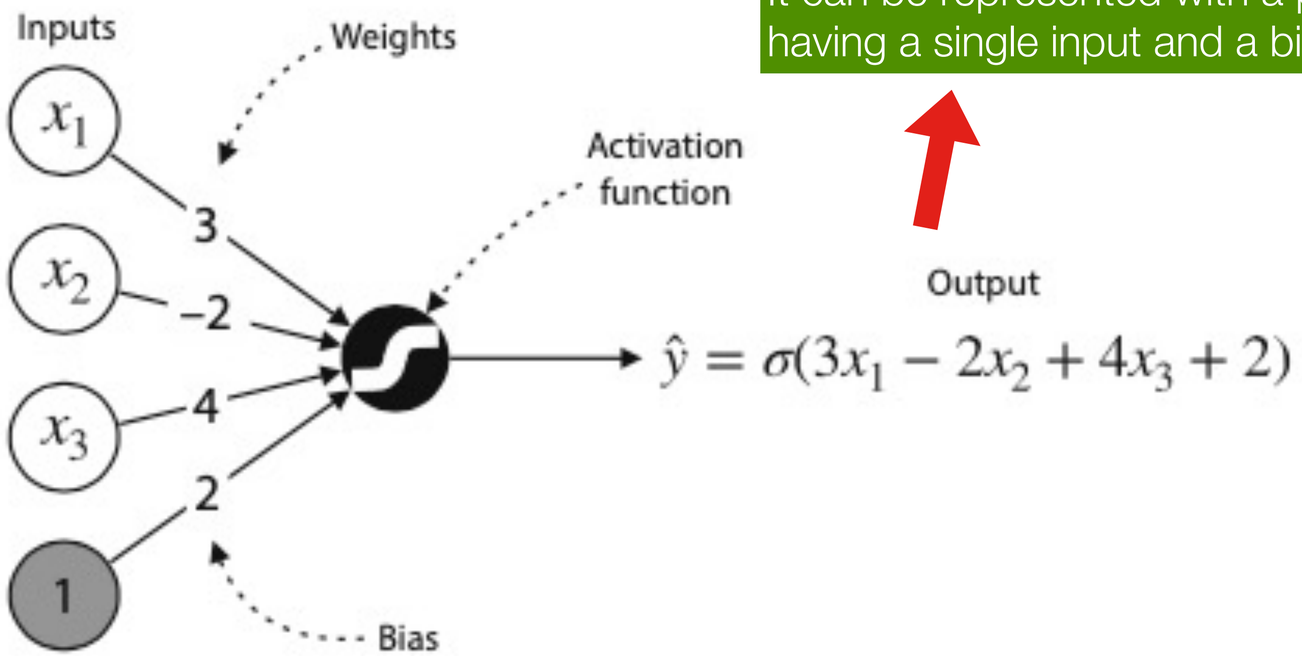
Perceptron


Inputs
Weights
Activation function
Output
x_1
x_2
x_3
1
3
−2
4
2
Bias
$\hat{y} = \sigma(3x_1 - 2x_2 + 4x_3 + 2)$

Perceptron

Cost = *b* + *wSize**

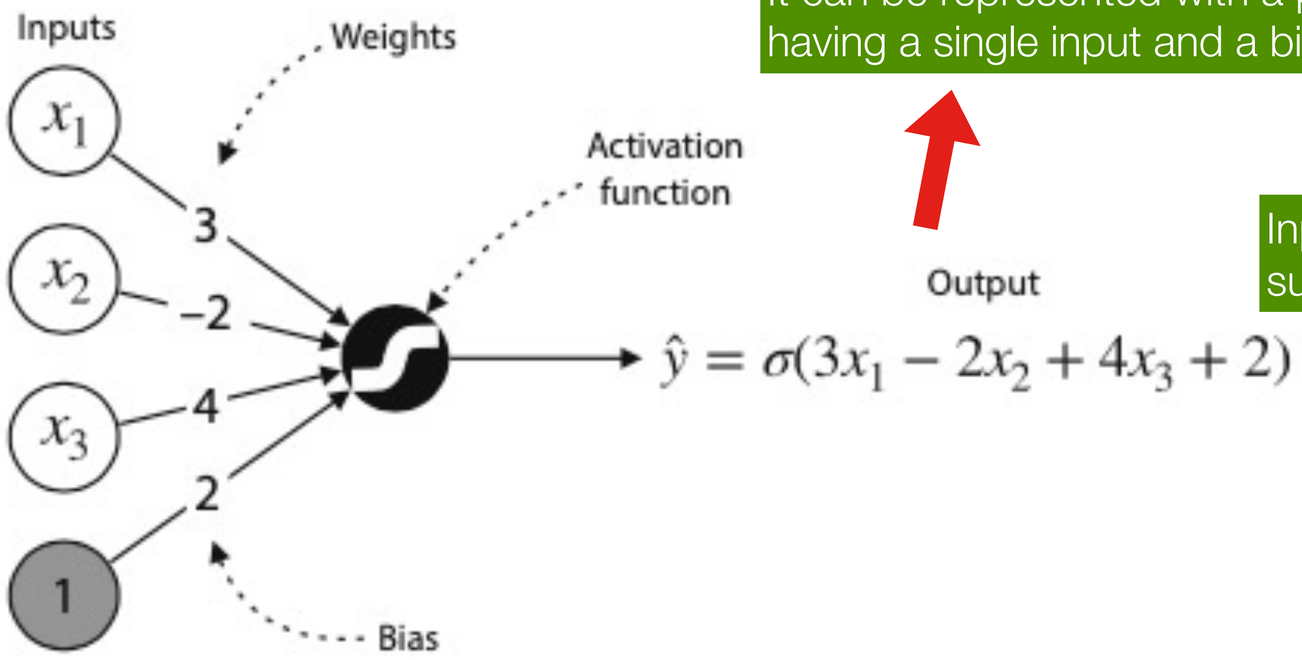
Remember our linear regression function?
It can be represented with a perceptron having a single input and a bias

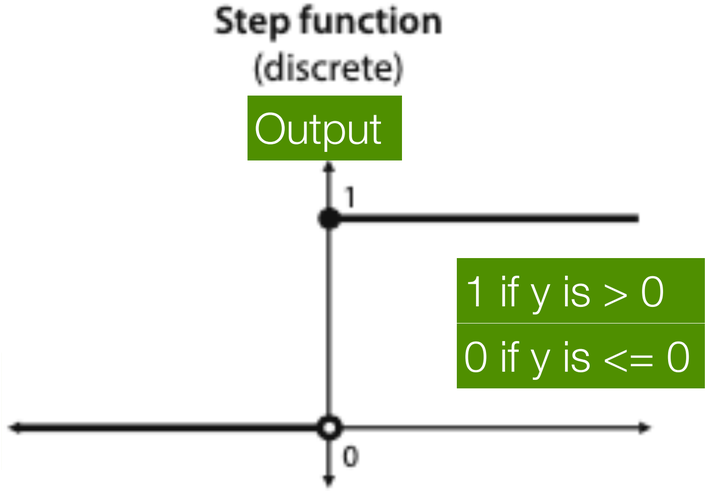


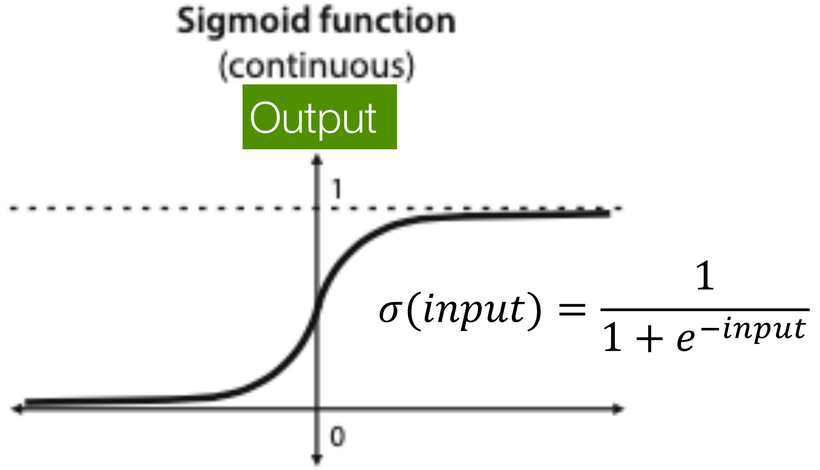
Perceptron

Cost = *a* + *wSize**

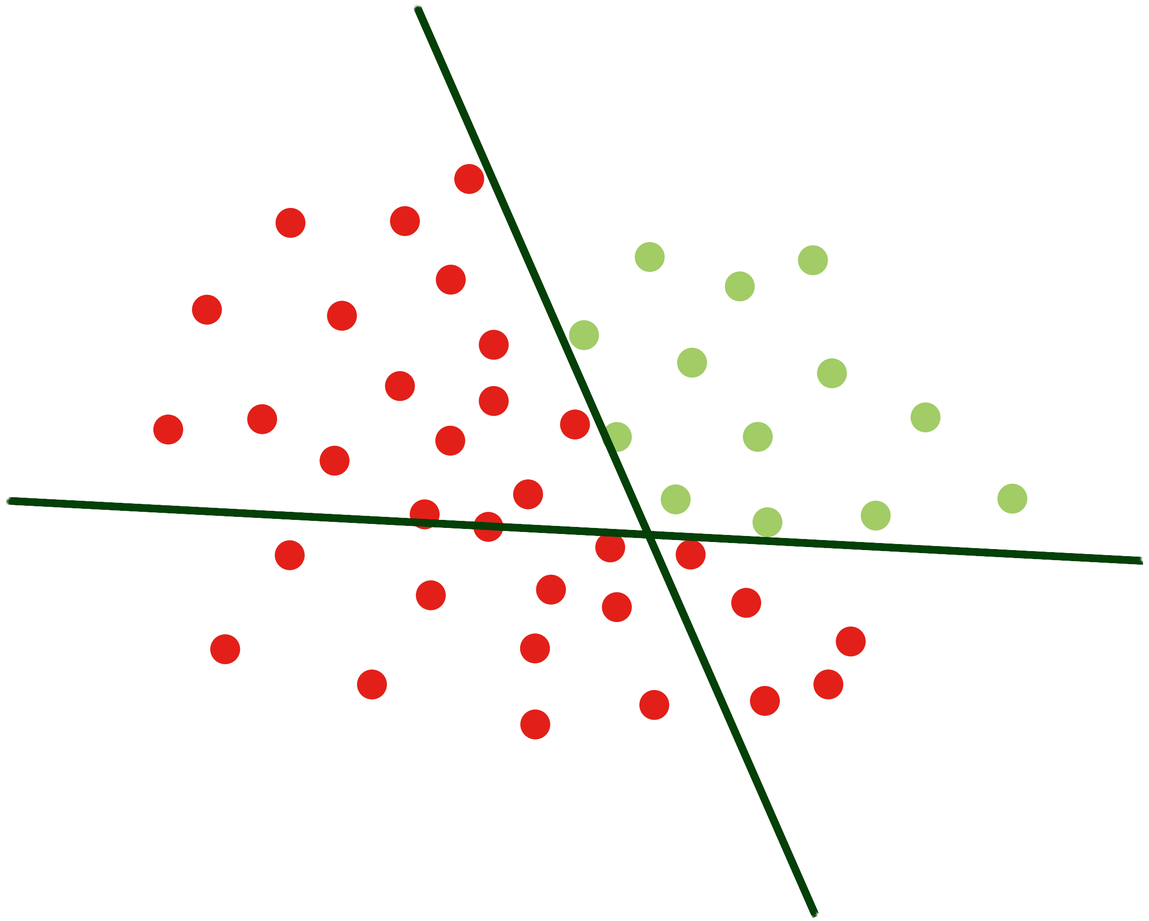
Remember our linear regression function? It can be represented with a perceptron having a single input and a bias



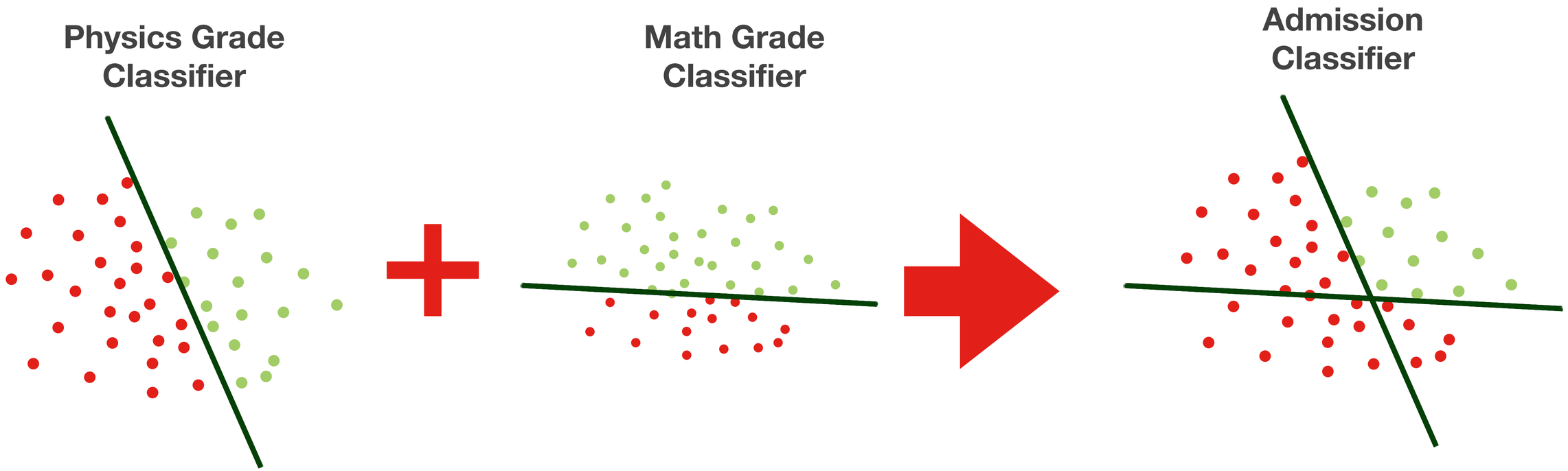




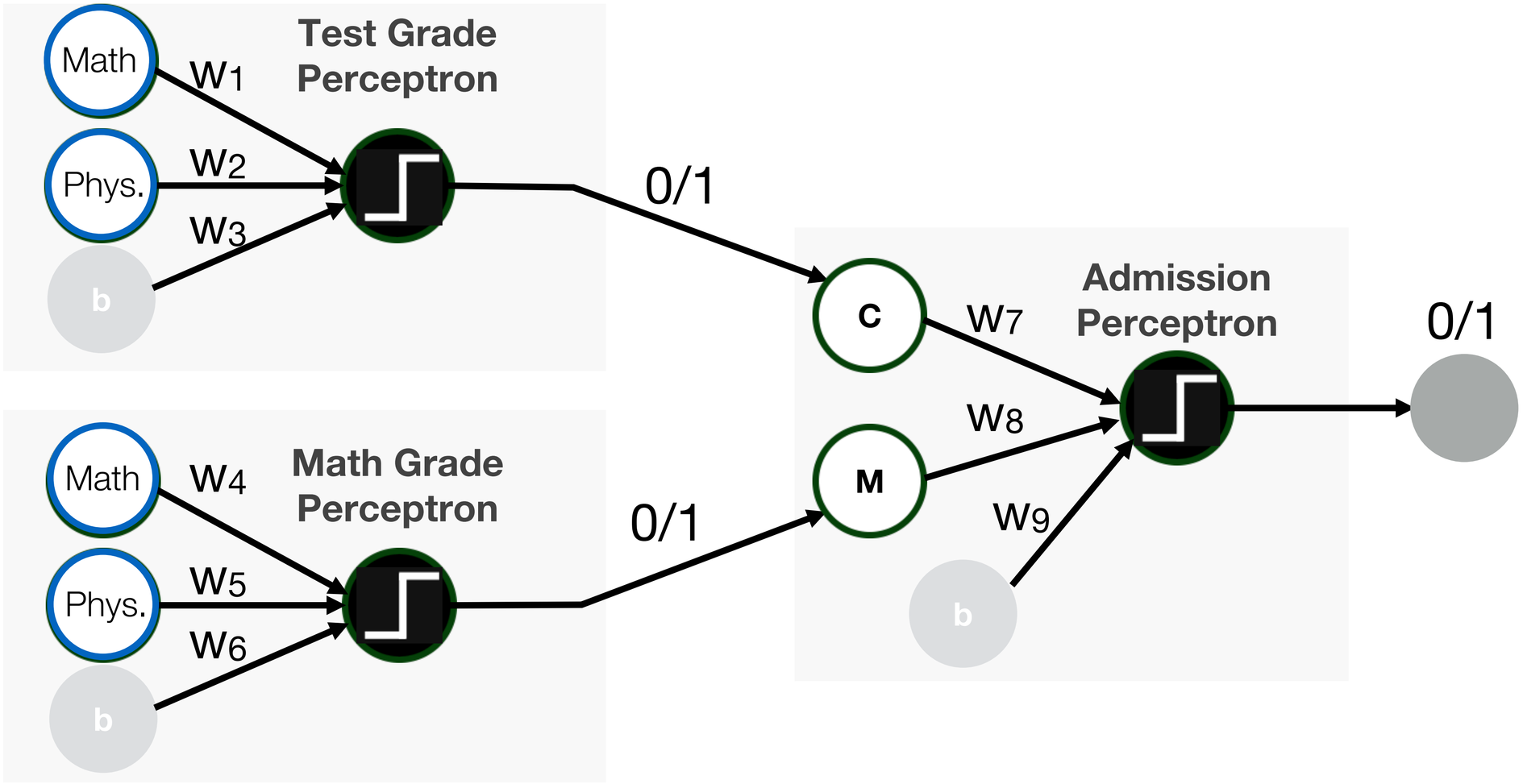
Neural Network



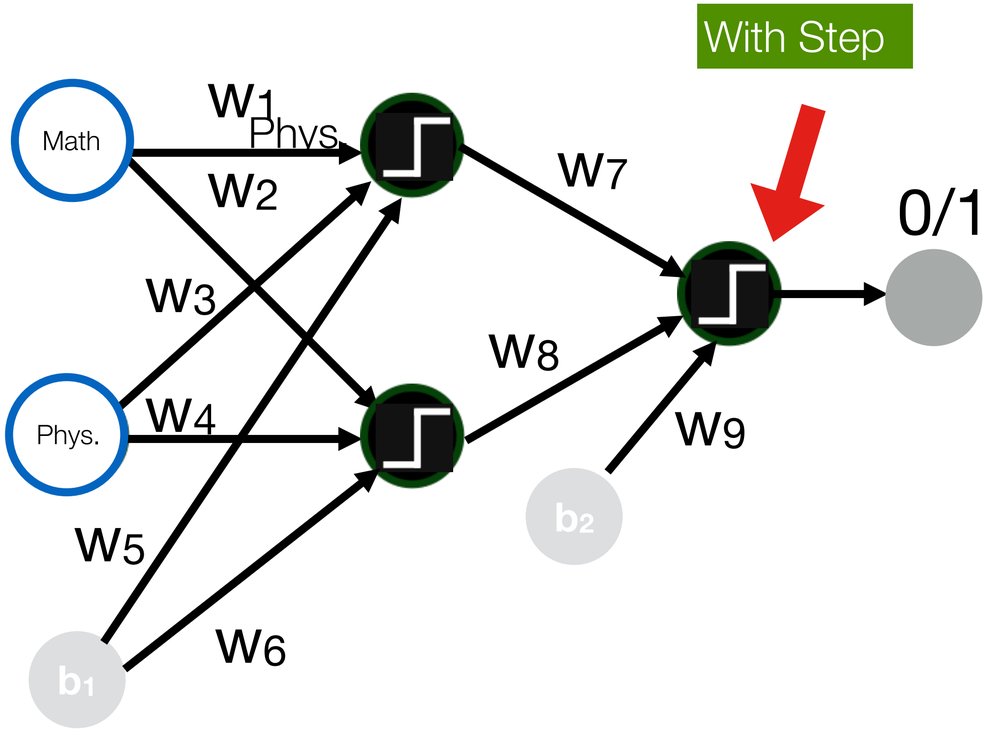
Neural Network

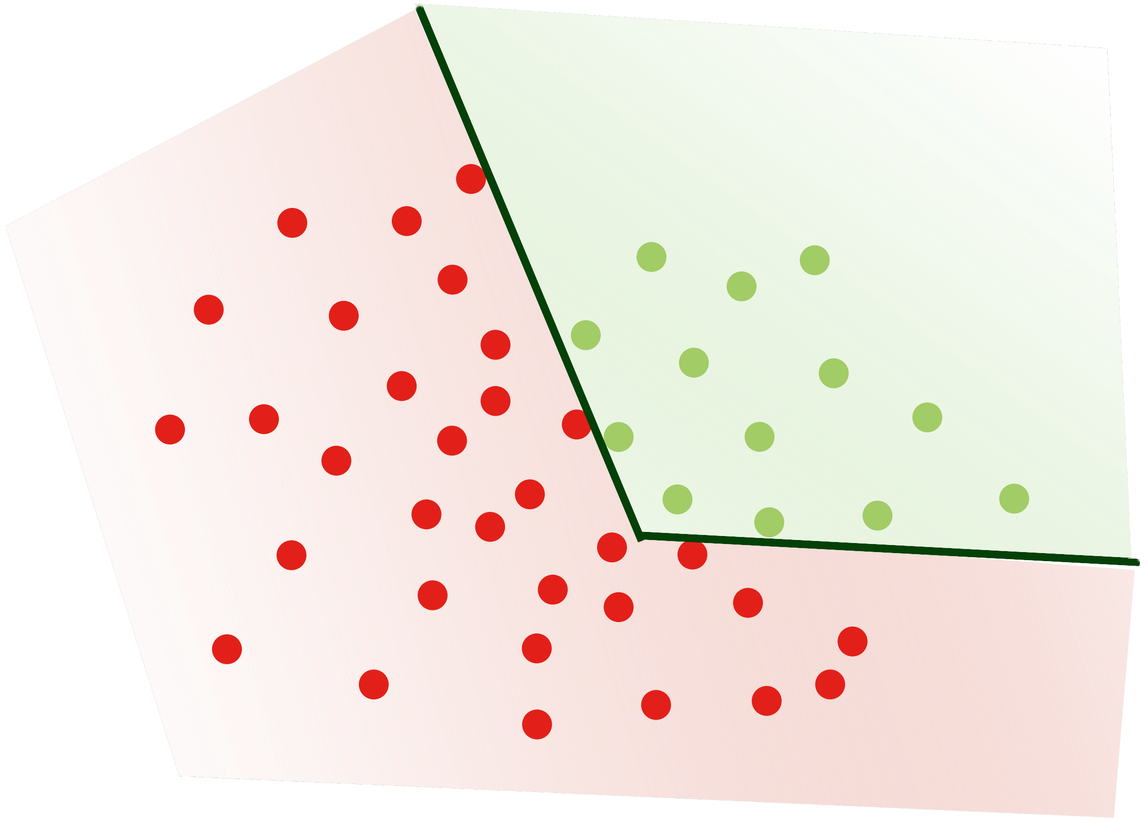


Neural Network

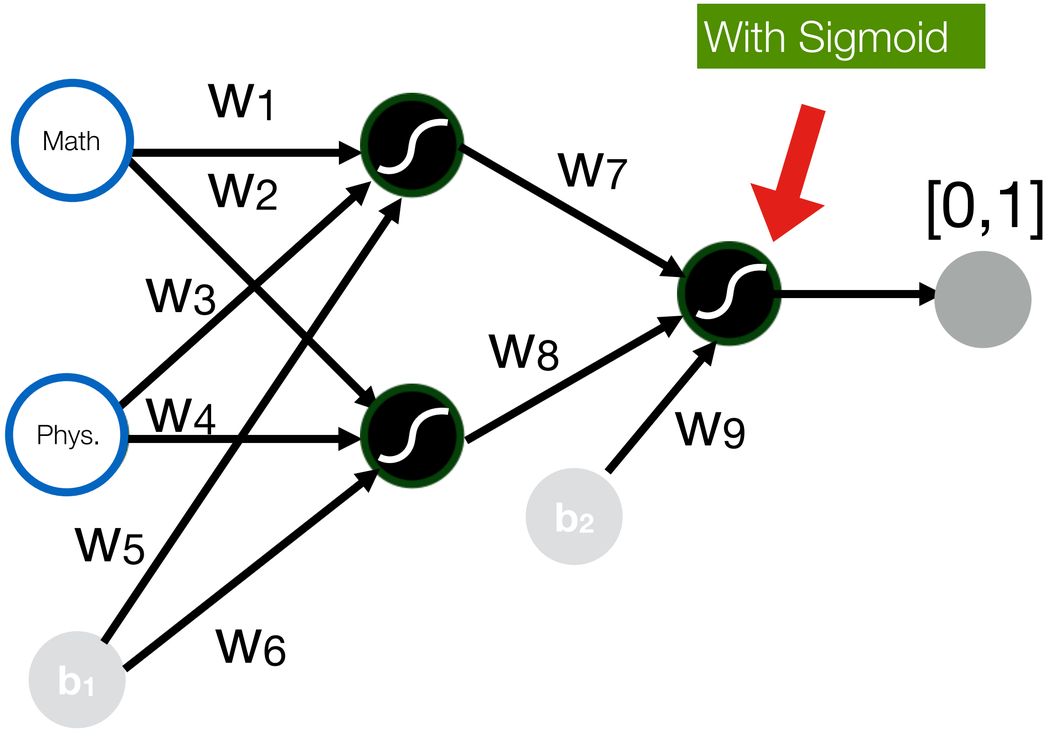


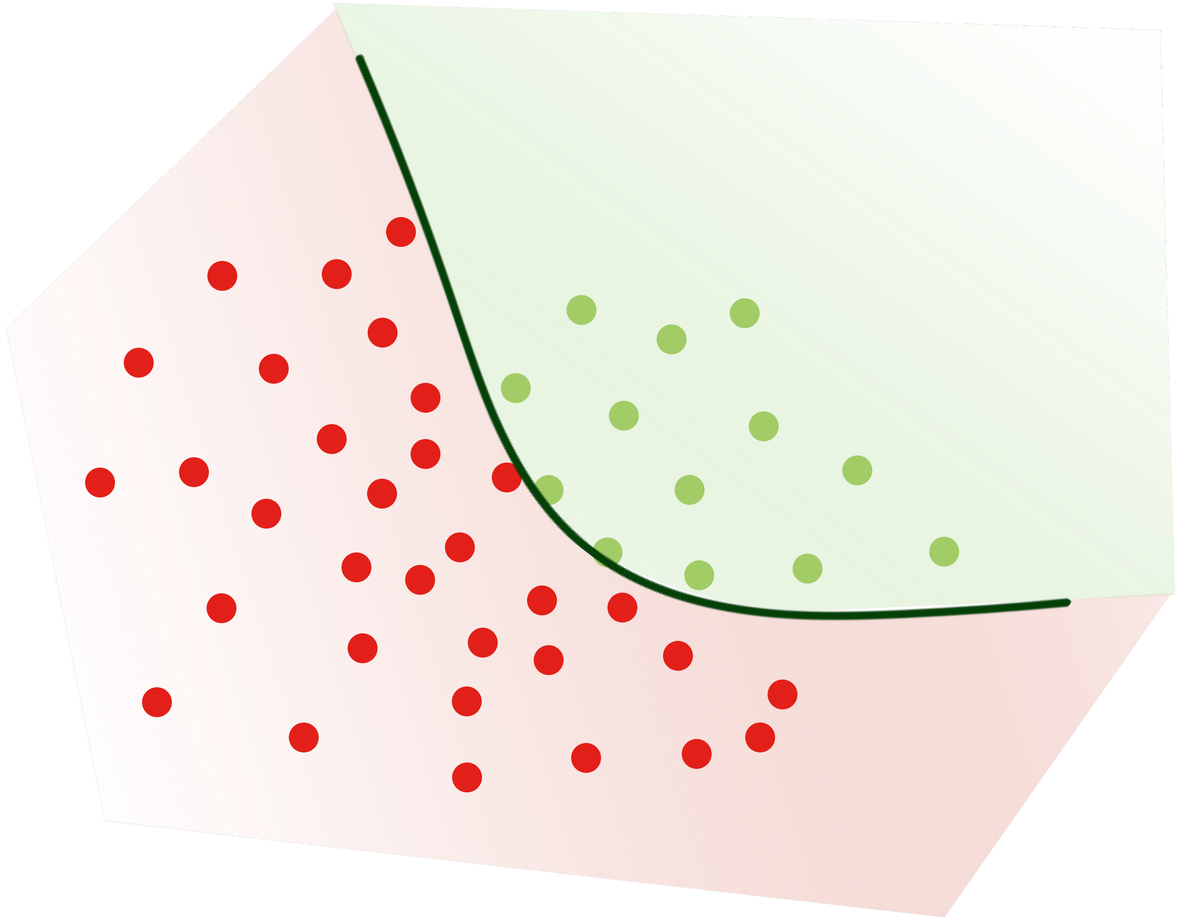
Neural Network



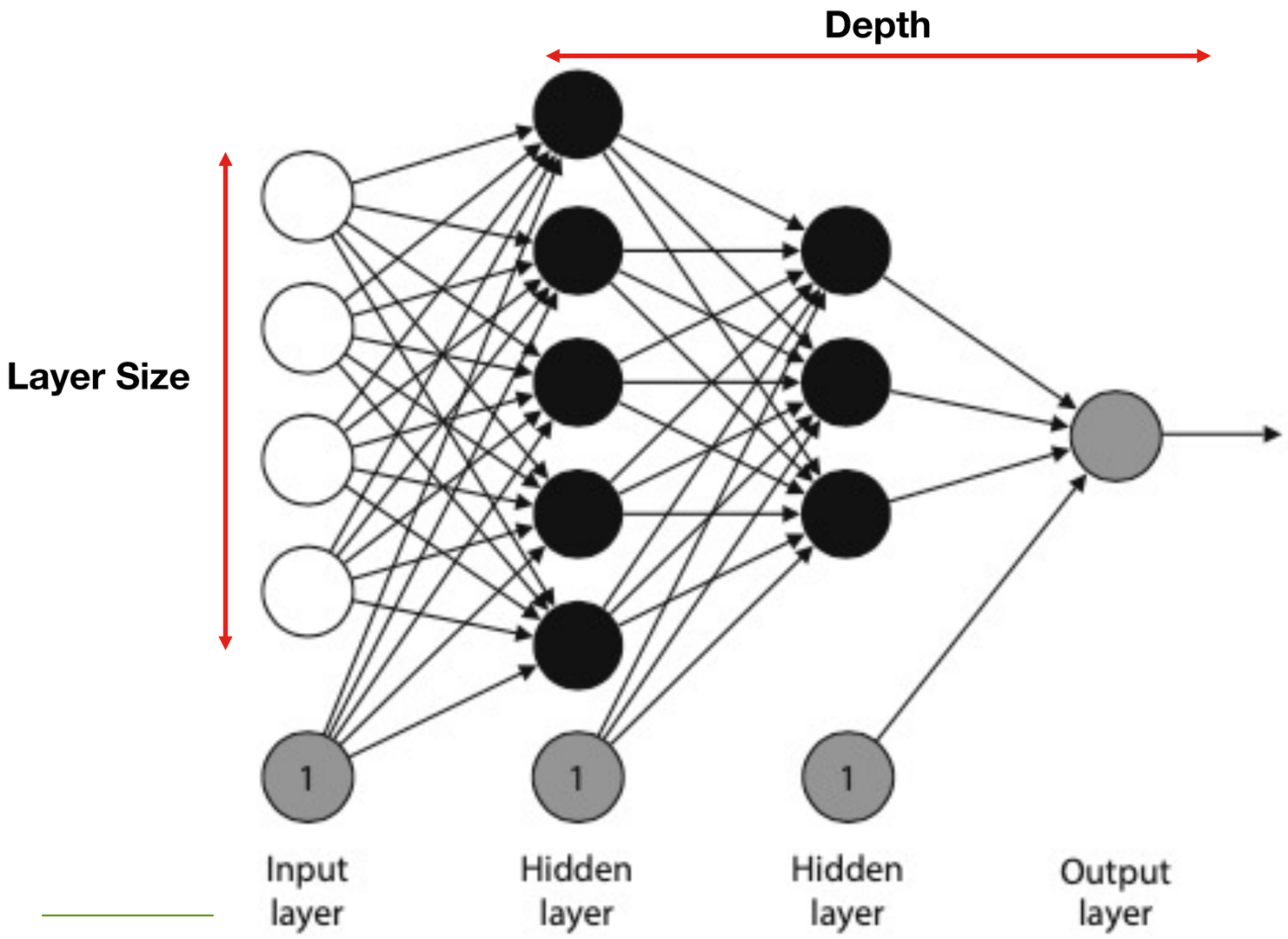


Neural Network





Fully Conected Neural Network

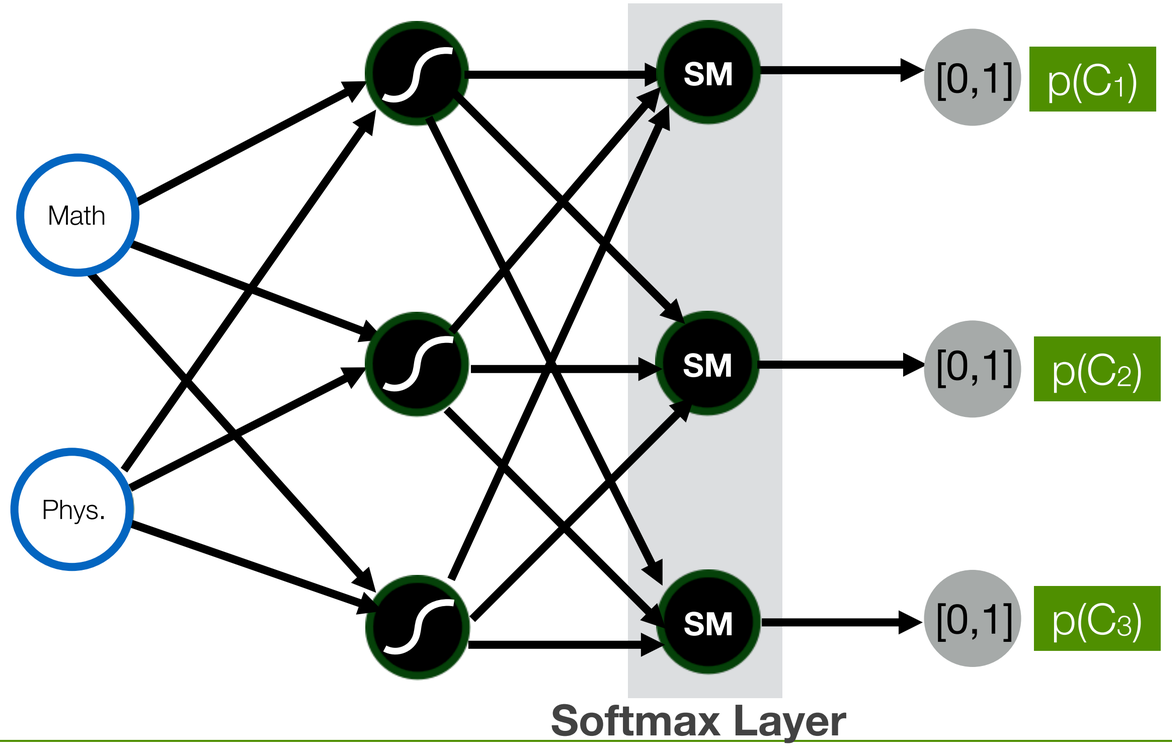


- **Hyperparameters**
  - Learning rate
  - Number of epochs
  - Architecture
    - # layers, #nodes, activation functions
  - Batch vs. mini-batch vs. stochastic gradient descent
  - Regularization parameters:
  - Dropout probability p

Advanced → next lecture

Classifying into multiple classes - softmax function

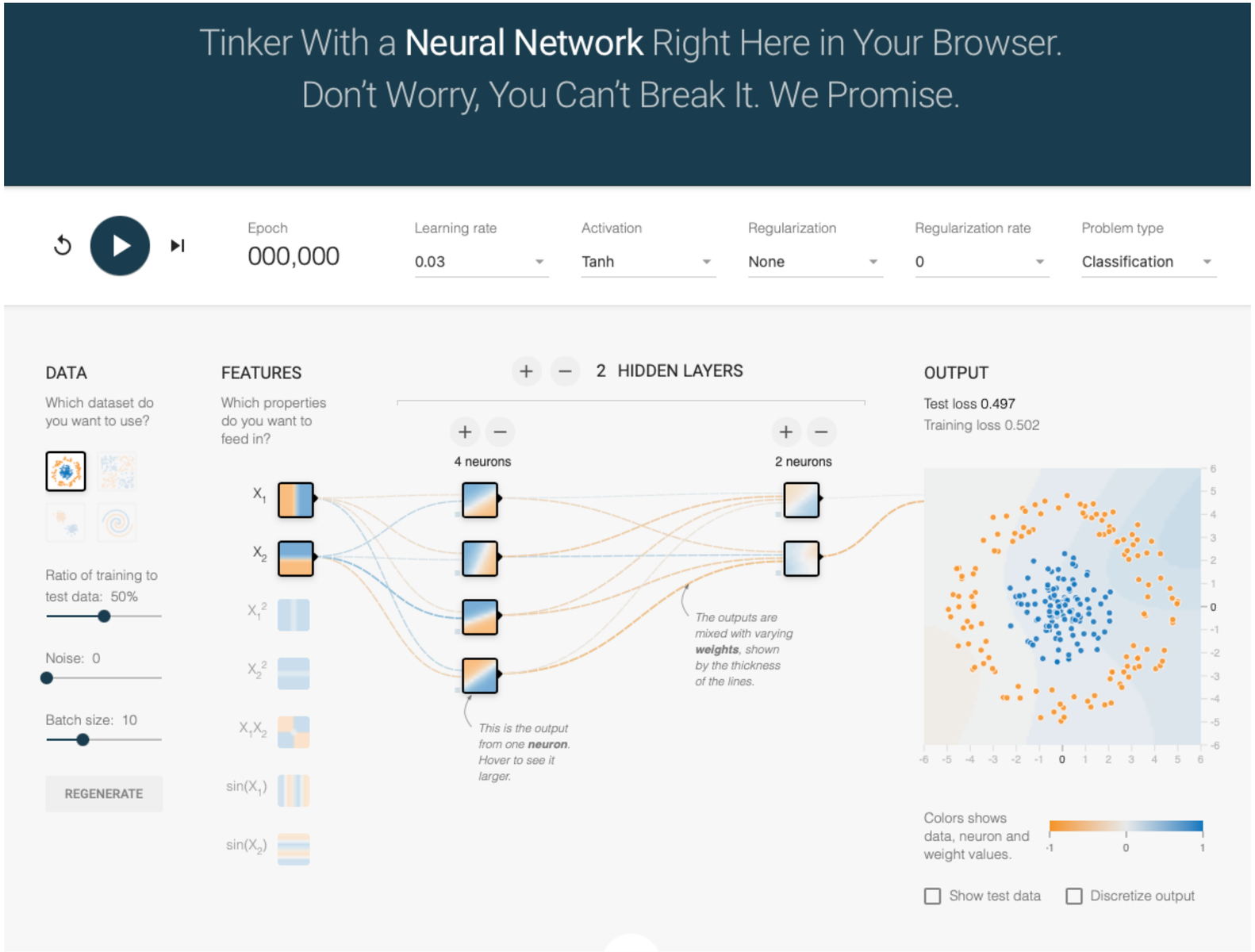
- Return a probability for **each class**
  - Imagine example C_1= *ADMITTED*, C_2 = *NOT ADMITTED*, C_3 = *TAKE TEST*
  - $p(C_1) = 0.37$, $p(C_2) = 0,21$, $p(C_3) = 0,42$
- We use the Softmax activation function for the output layer



$$Softmax(x_i) = \frac{e^{(x_i)}}{\sum_j^K e^{(x_j)}}$$

Value of class *i*

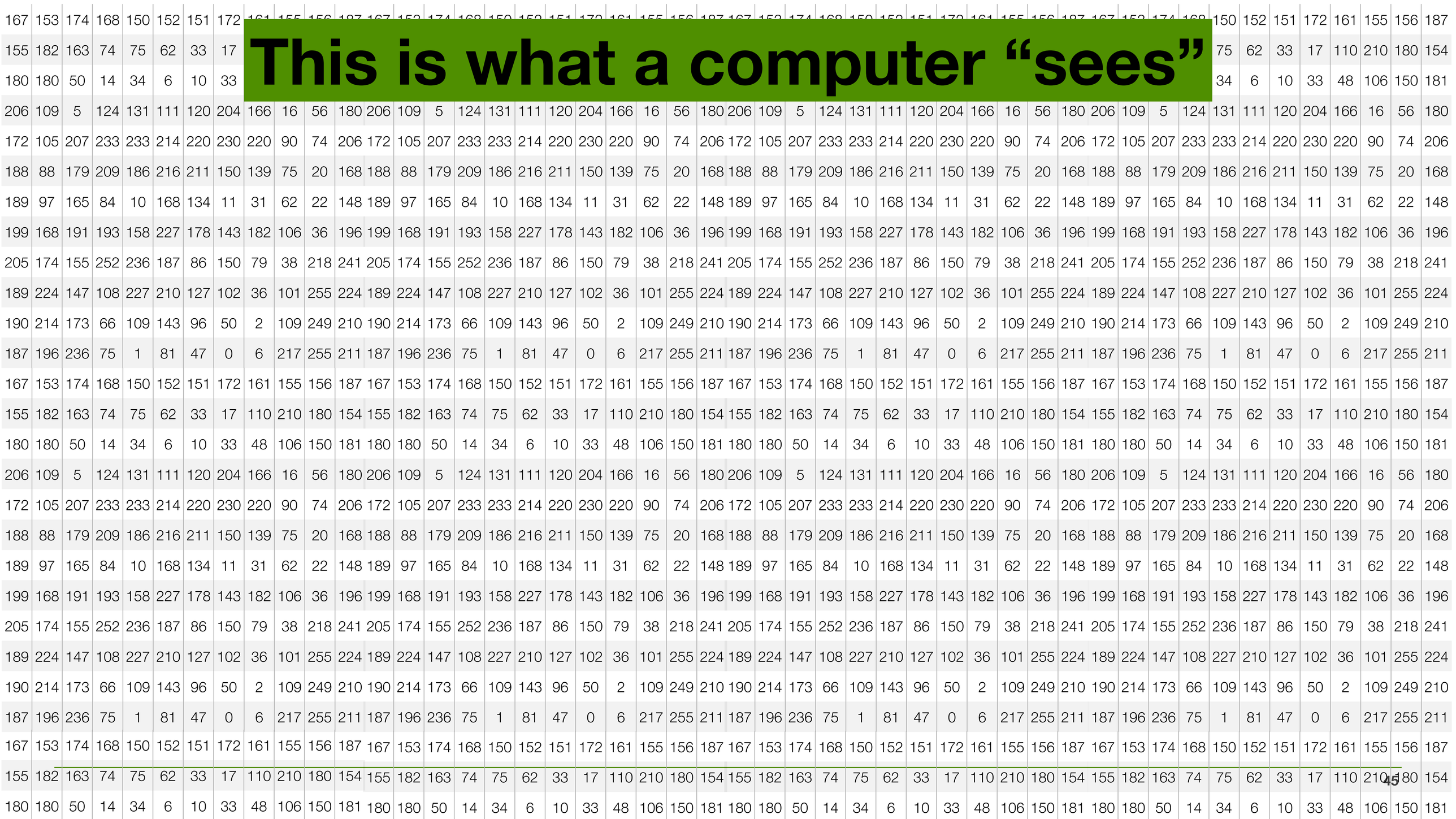
Normalisation term on K classes



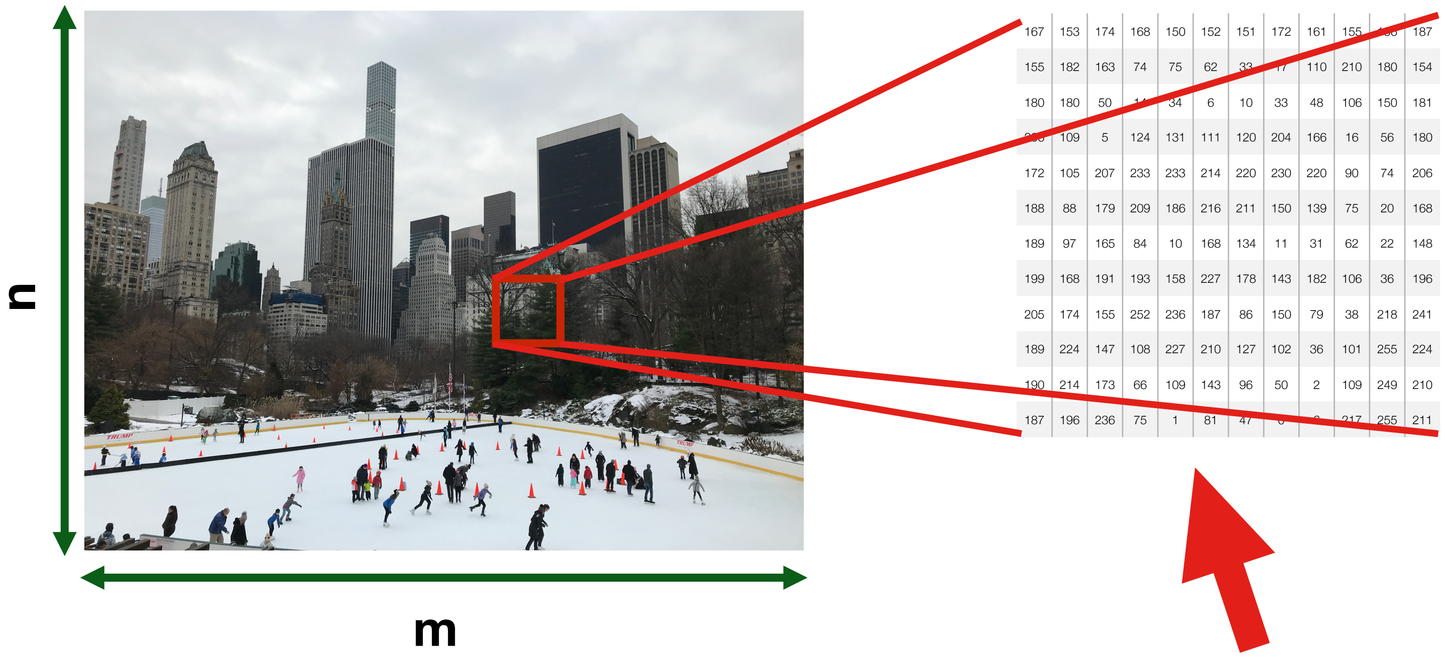
https://playground.tensorflow.org/

Machine Learning and Images


What do you see?
TRUMP
TRUMP


This is what a computer “sees”

Images



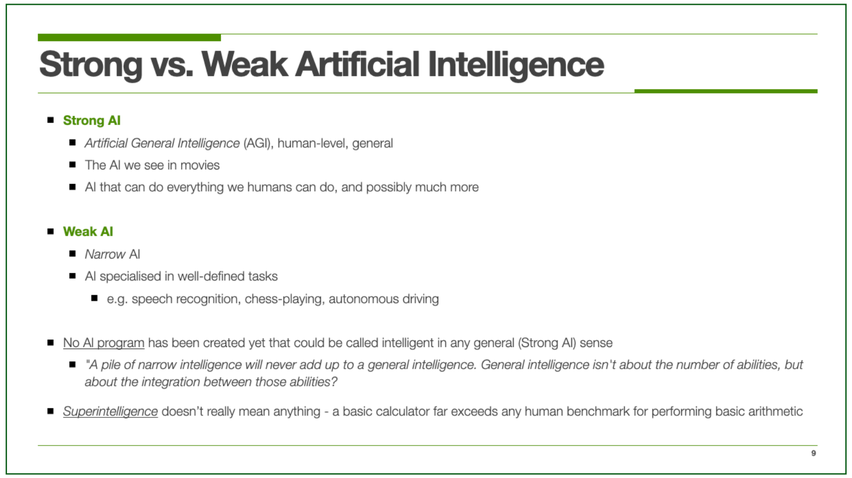
- Each pixel in an image is a feature
- Dimensionality
  - n x m
- Feature (pixel) values are numerical
  - 0 or 1 for Black and White
  - Between 0 and 255 for greyscale
  - 16M values for RGB

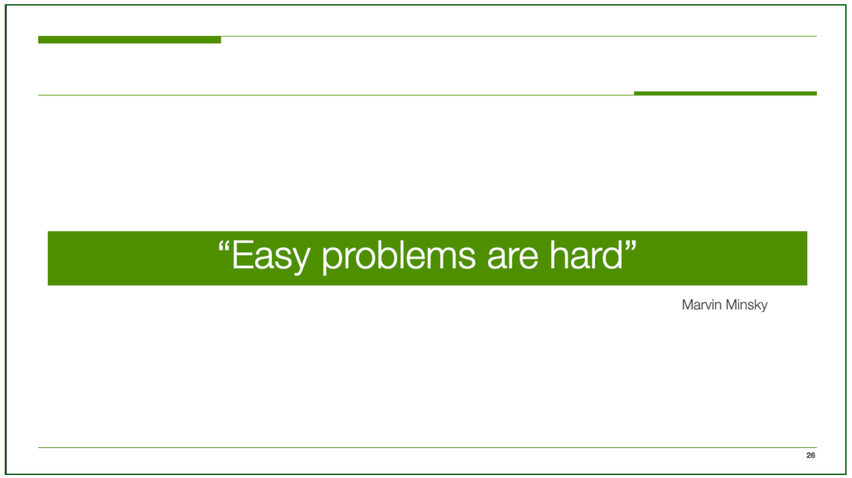
Computer Vision

- Building algorithms that can “understand” the content of images and use it for other applications
- It is a “Strong AI” problem
  - signal-to-symbol conversion
  - The **semantic gap**

- A general-purpose vision system requires
  - Flexible, robust visual representatation
    - Updated and maintaned
  - Reasoning
  - Interfacing with attention goals, and plans

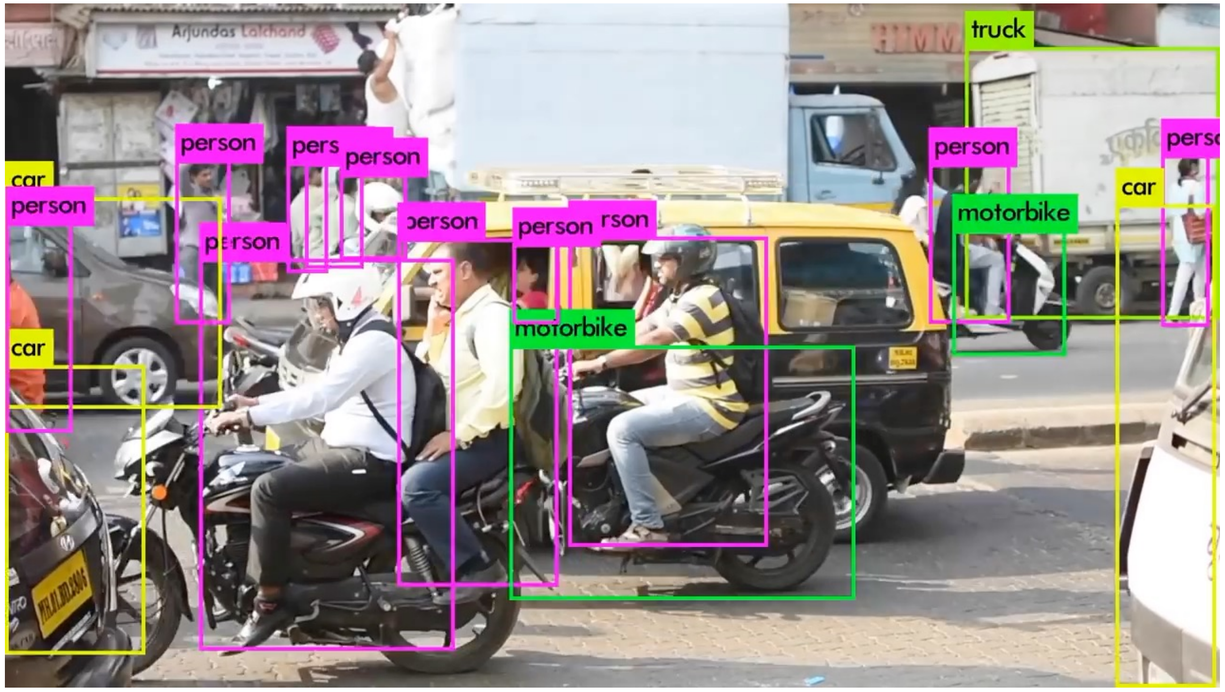
- What specific tasks can we train a CV system to perform?

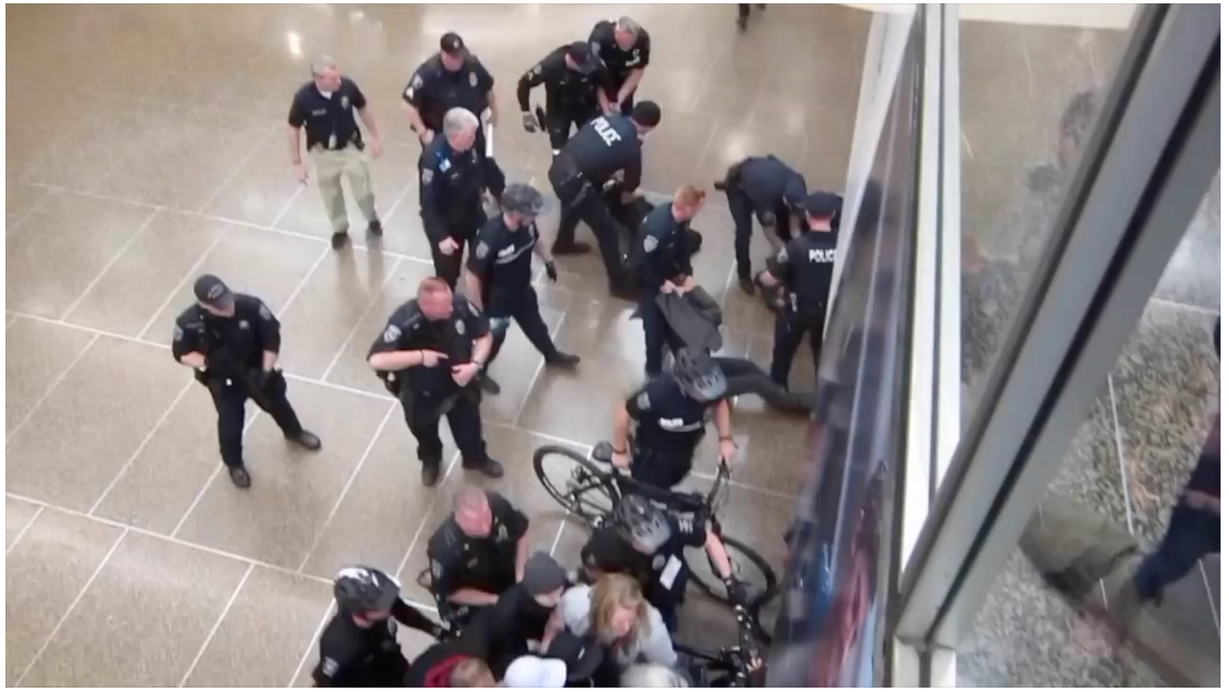


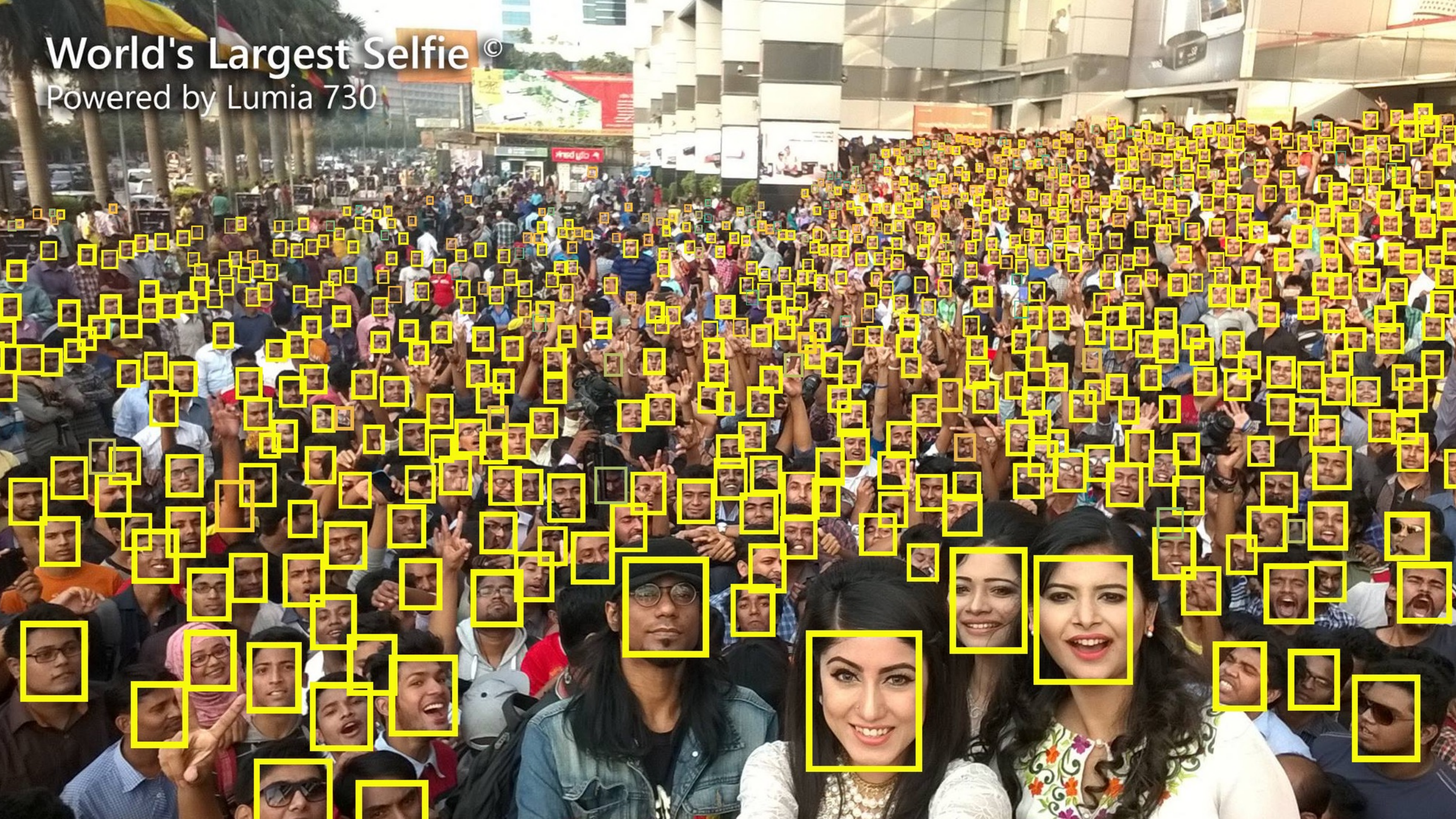


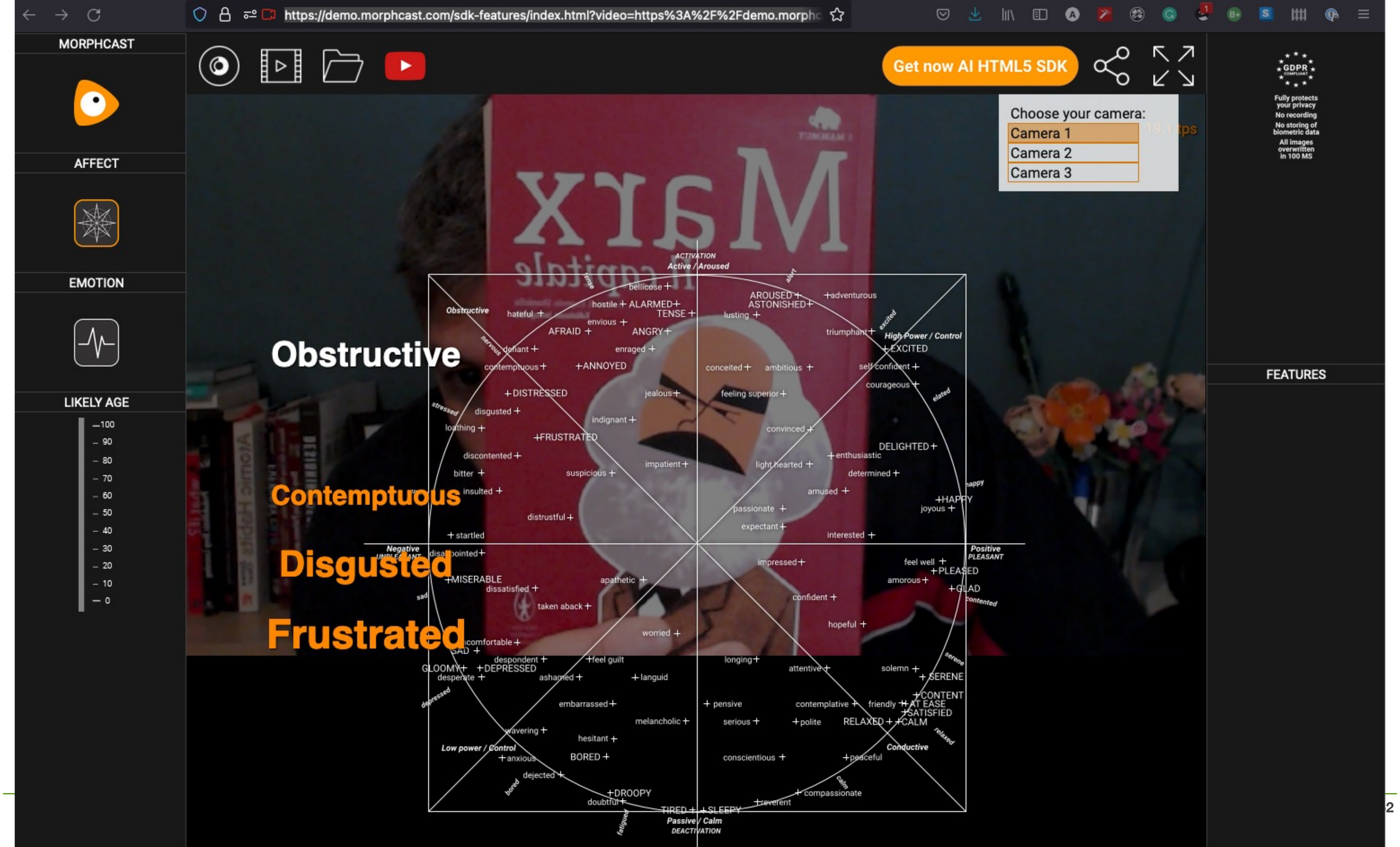

Is this a flag?
(Recognition / Classification)
TRUMP
TRUMP


Where are the people?
(Recognition/Classification)
TRUMP
TRUMP

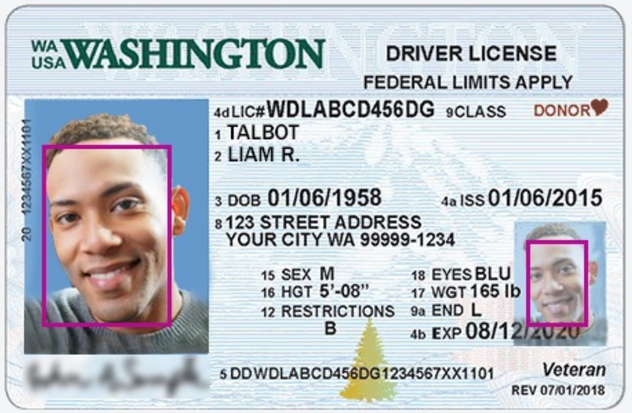

truck
person
person
person
car
person
person
motorbike
car
motorbike




World's Largest Selfie ©
Powered by Lumia 730

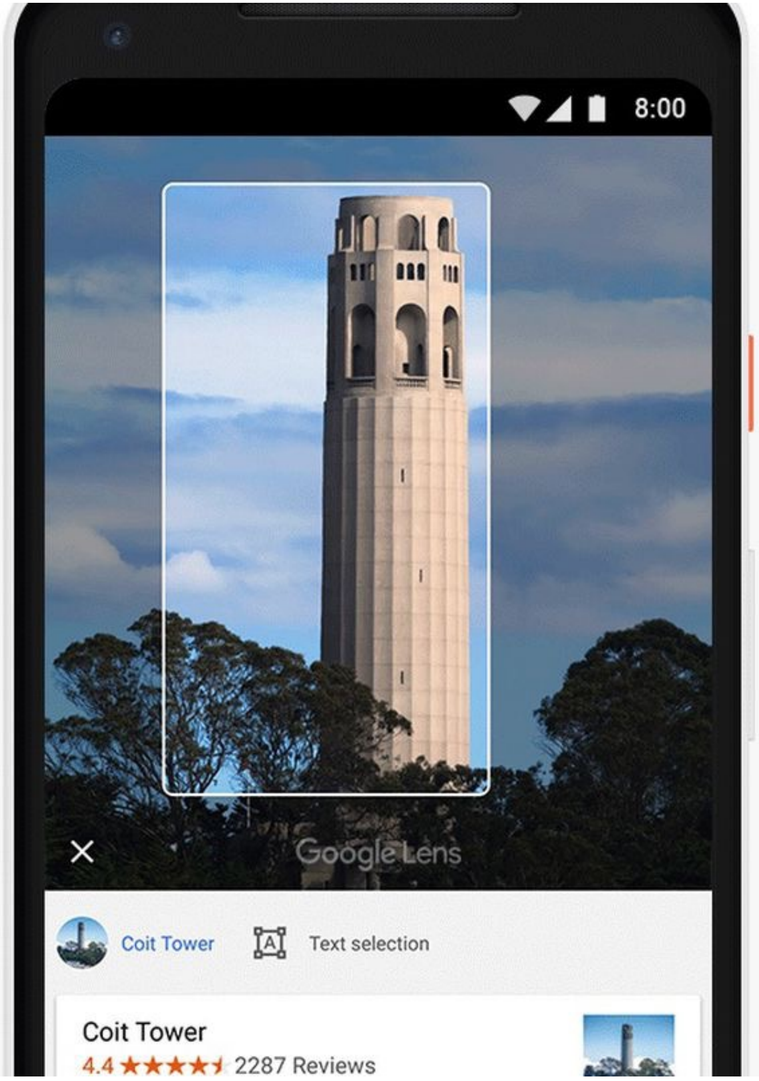

https://demo.morphcast.com/sdk-features/index.html?video=https%3A%2F%2Fdemo.morphc
MORPHCAST
AFFECT
EMOTION
LIKELY AGE
100
90
80
70
60
50
40
30
20
10
0
Get now AI HTML5 SDK
Choose your camera:
Camera 1
Camera 2
Camera 3
GDPR COMPLIANT
Fully protects your privacy
No recording
No storing of biometric data
All images overwritten in 100 MS
FEATURES
Obstructive
Contemptuous
Disgusted
Frustrated
ACTIVATION
Active / Aroused
Passive / Calm
DEACTIVATION
Negative
Positive
PLEASANT
High Power / Control
Low power / Control
Conductive
Obstructive
bellicose
hostile
ALARMED
hateful
TENSE
envious
AFRAID
ANGRY
defiant
enraged
contemptuous
ANNOYED
DISTRESSED
jealous
disgusted
indignant
loathing
FRUSTRATED
discontented
bitter
suspicious
impatient
insulted
distrustful
startled
disappointed
MISERABLE
dissatisfied
apathetic
taken aback
worried
uncomfortable
SAD
despondent
feel guilt
GLOOMY
DEPRESSED
desperate
ashamed
languid
embarrassed
melancholic
wavering
hesitant
Low power / Control
anxious
BORED
dejected
DROOPY
doubtful
TIRED
SLEEPY
AROUSED
ASTONISHED
adventurous
lusting
triumphant
EXCITED
conceited
ambitious
self confident
courageous
feeling superior
convinced
DELIGHTED
enthusiastic
light hearted
determined
amused
passionate
HAPPY
joyous
expectant
interested
impressed
feel well
PLEASED
amorous
GLAD
confident
hopeful
longing
attentive
solemn
SERENE
CONTENT
pensive
contemplative
friendly
AT EASE
SATISFIED
serious
polite
RELAXED
CALM
conscientious
peaceful
compassionate
reverent
alert
excited
elated
happy
contented
serene
relaxed
calm
bored
fatigued
nervous
stressed
sad
depressed

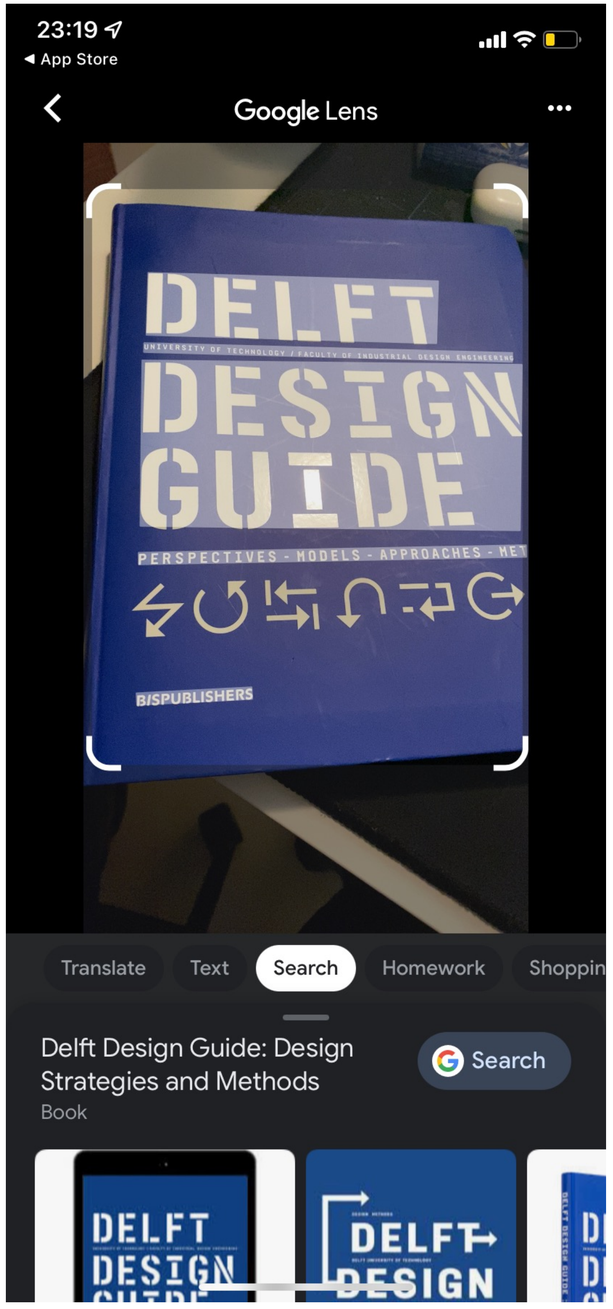

Is this Jeff?
(Identification)
TRUMP
TRUMP

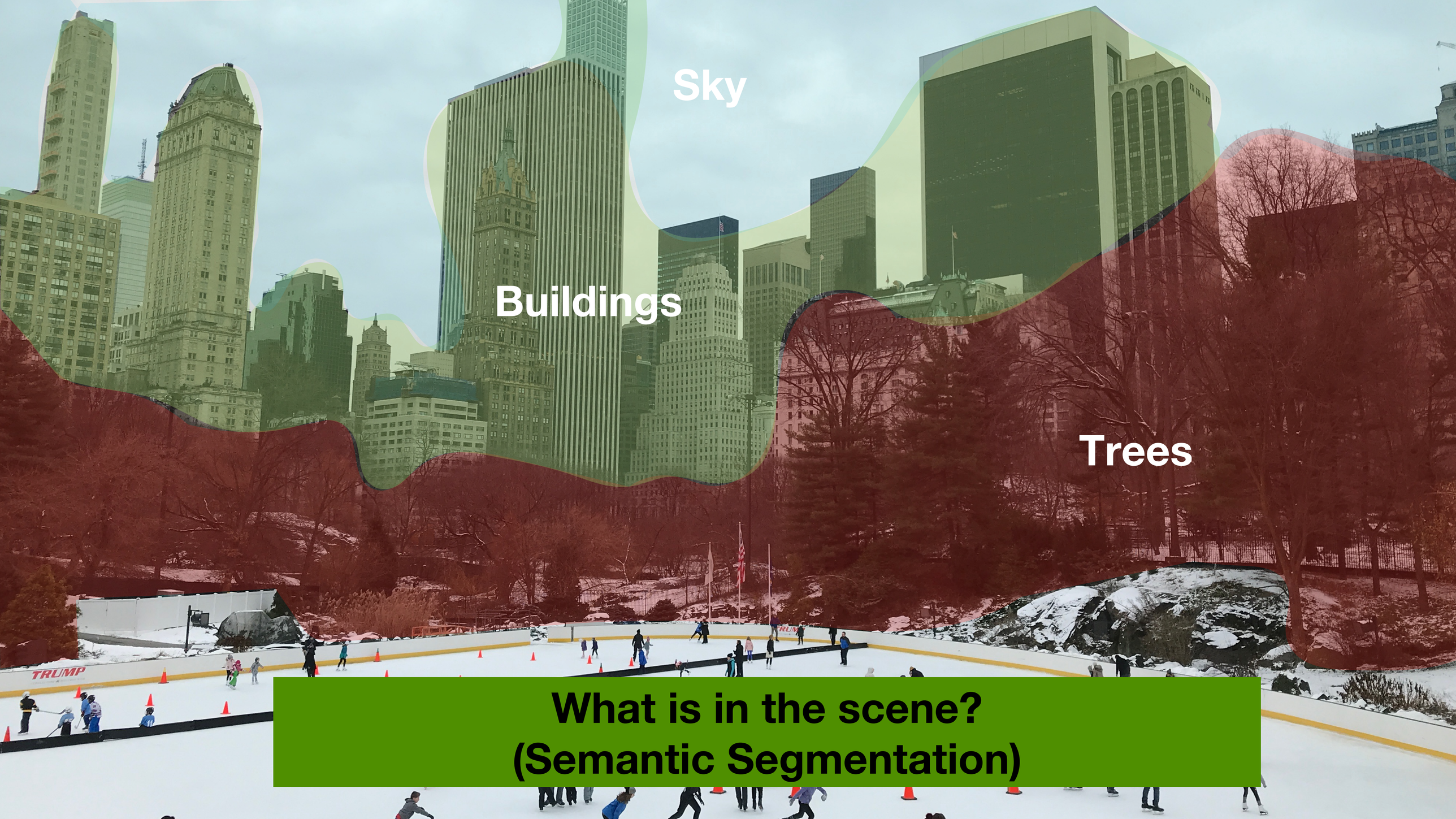

WA
USA
WASHINGTON
DRIVER LICENSE
FEDERAL LIMITS APPLY
4d LIC# WDLABCD456DG
9 CLASS
DONOR
1 TALBOT
2 LIAM R.
3 DOB 01/06/1958
4a ISS 01/06/2015
8 123 STREET ADDRESS
YOUR CITY WA 99999-1234
15 SEX M
16 HGT 5'-08"
12 RESTRICTIONS B
18 EYES BLU
17 WGT 165 lb
9a END L
4b EXP 08/12/2020
5 DDWDLABCD456DG1234567XX1101
Veteran
REV 07/01/2018


撮影画像
登録済画像
対象人物
撮影画像
登録済画像
対象人物
撮影画像
登録済画像
対象人物
撮影画像
登録済画像
対象人物

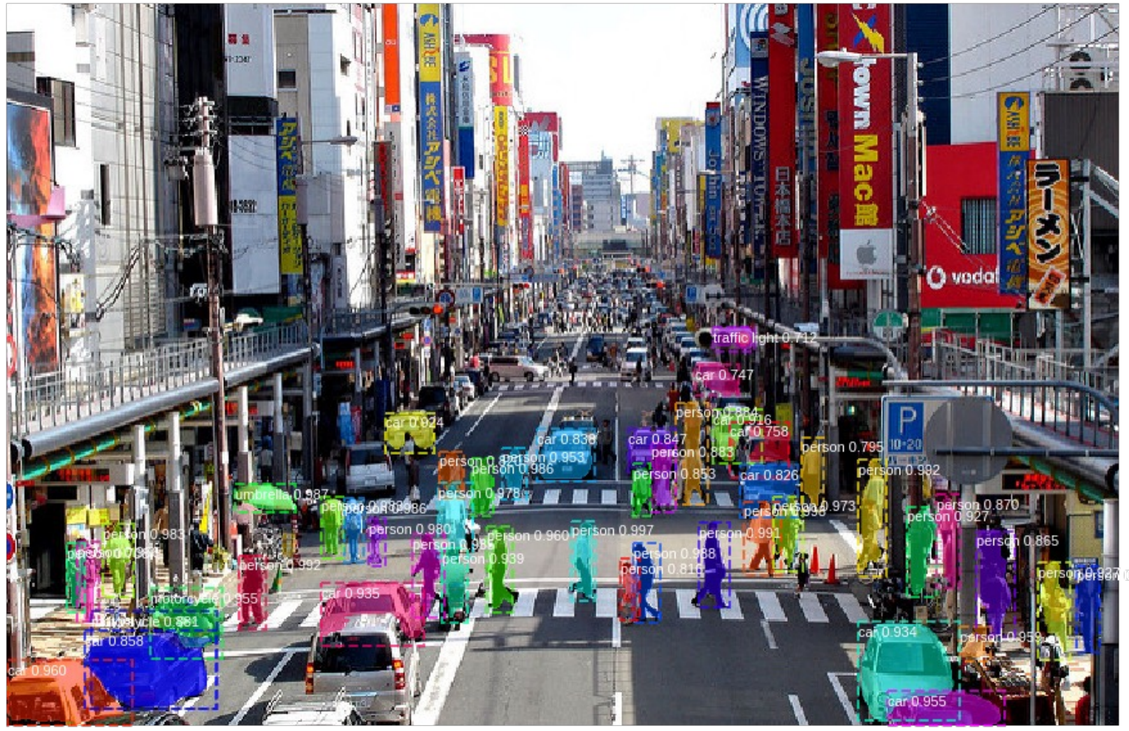

Is this the Wollman Rink?
(Identification)
TRUMP
TRUMP

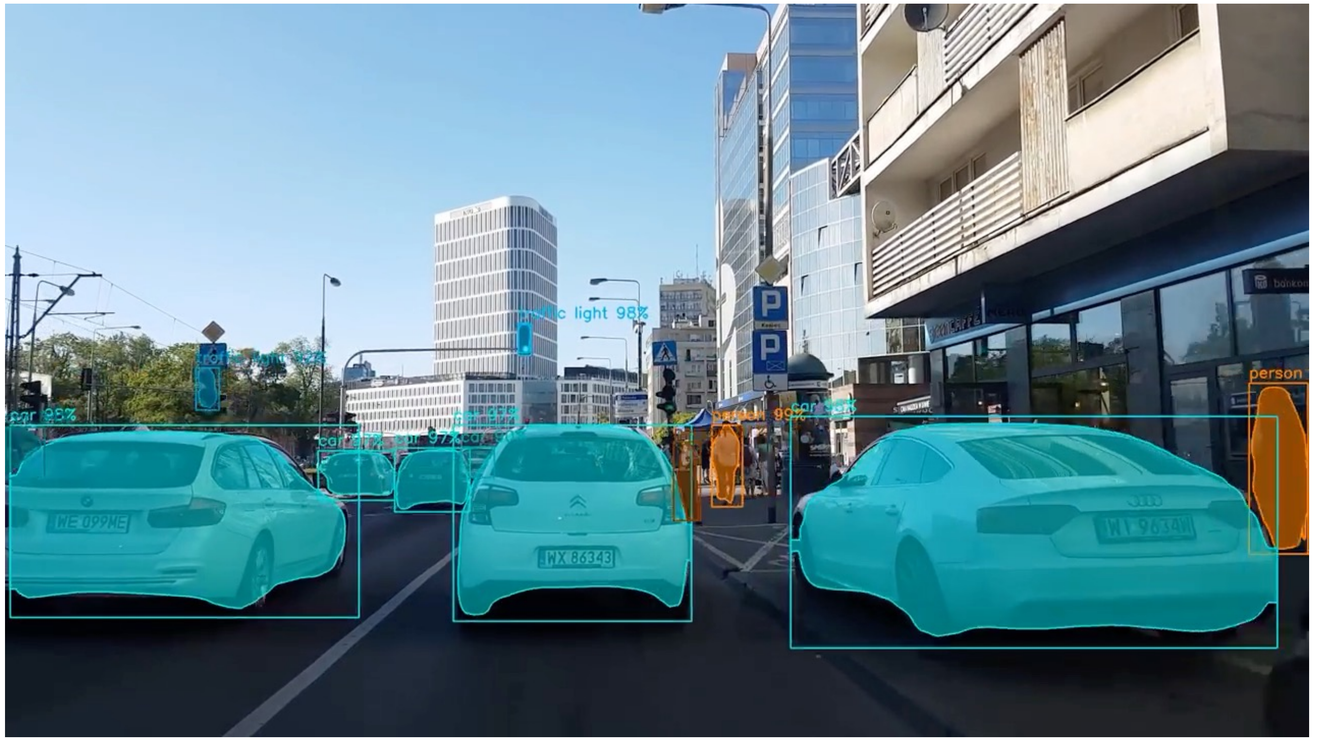

8:00
Google Lens
Coit Tower
Text selection
Coit Tower
4.4 2287 Reviews


23:19
App Store
Google Lens
DELFT
DESIGN
GUIDE
PERSPECTIVES - MODELS - APPROACHES
BISPUBLISHERS
Translate
Text
Search
Homework
Shoppin
Delft Design Guide: Design Strategies and Methods
Book
Search



What is in the scene?
(Semantic Segmentation)

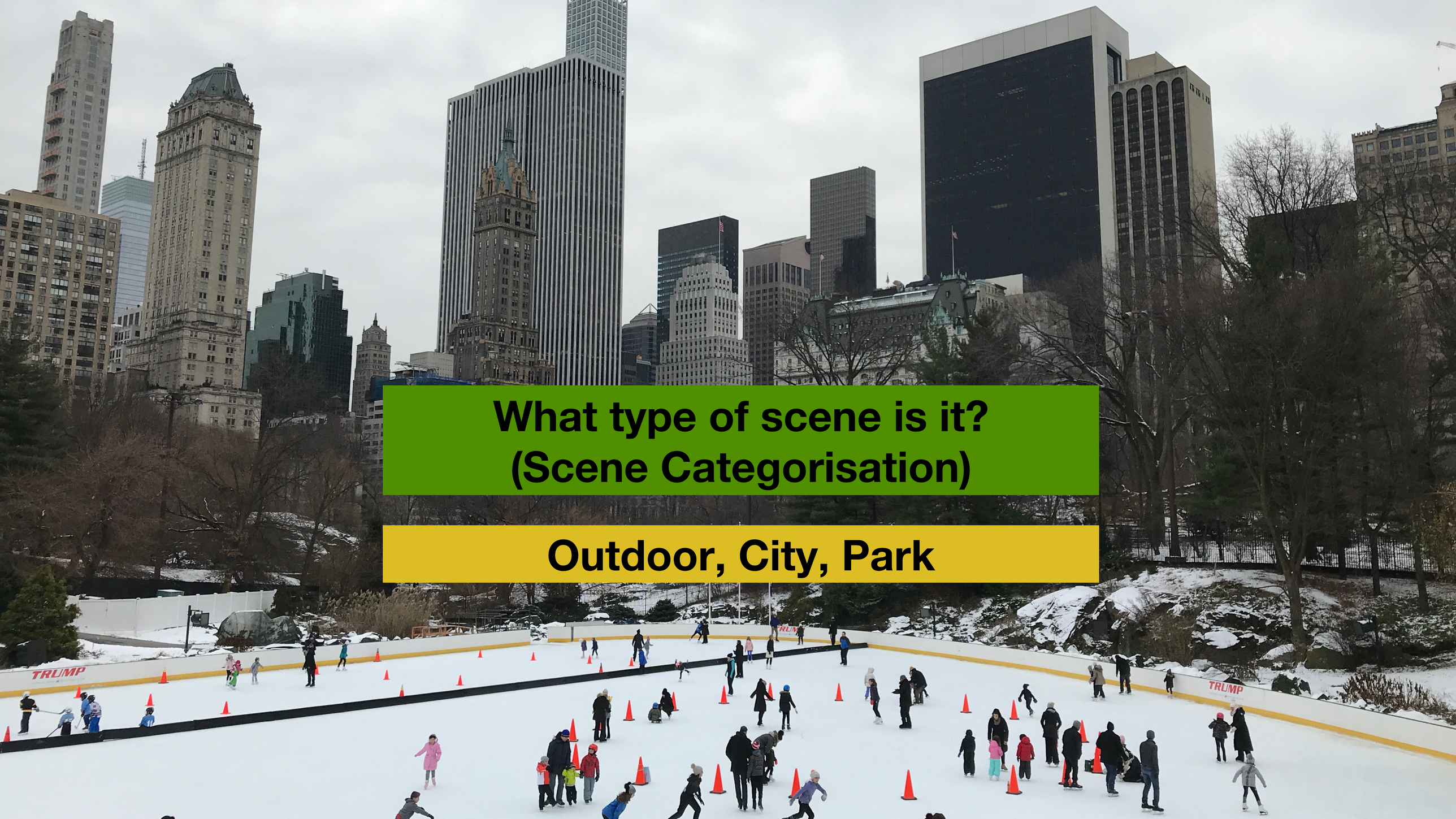




https://github.com/matterport/Mask_RCNN

Project Sunroof

https://www.google.com/get/sunroof


What type of scene is it?
(Scene Categorisation)
Outdoor, City, Park
TRUMP
TRUMP

Predictions:

- **Type of environment:** outdoor
- **Scene categories:** skyscraper (0.704), downtown (0.211)
- **Scene attributes:** man-made, vertical components, open area, natural light, clouds, no horizon, metal, glass, sunny
- **Informative region for predicting the category *skyscraper* is:**

Hugging Face | Search models, datasets, users... | Models | Datasets | Spaces | Docs | Solutions | Pricing

Spaces: OFA-Sys / **OFA-Image_Caption** | like 12 | Running

App | Files and versions

OFA-Image_Caption

Gradio Demo for OFA-Image_Caption. Upload your own image or click any one of the examples, and click "Submit" and then wait for the generated caption.

Image

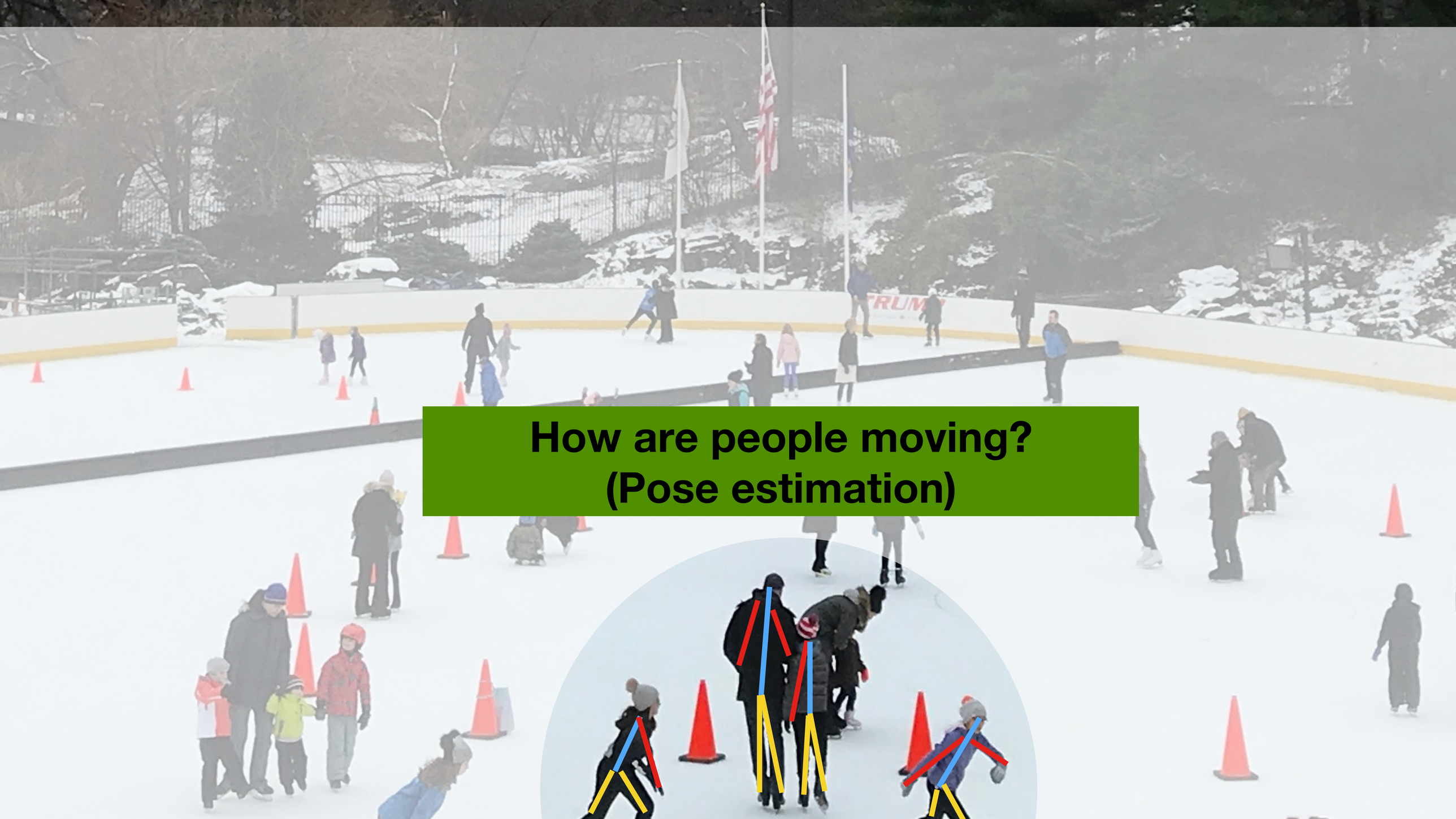
Caption 8.28s

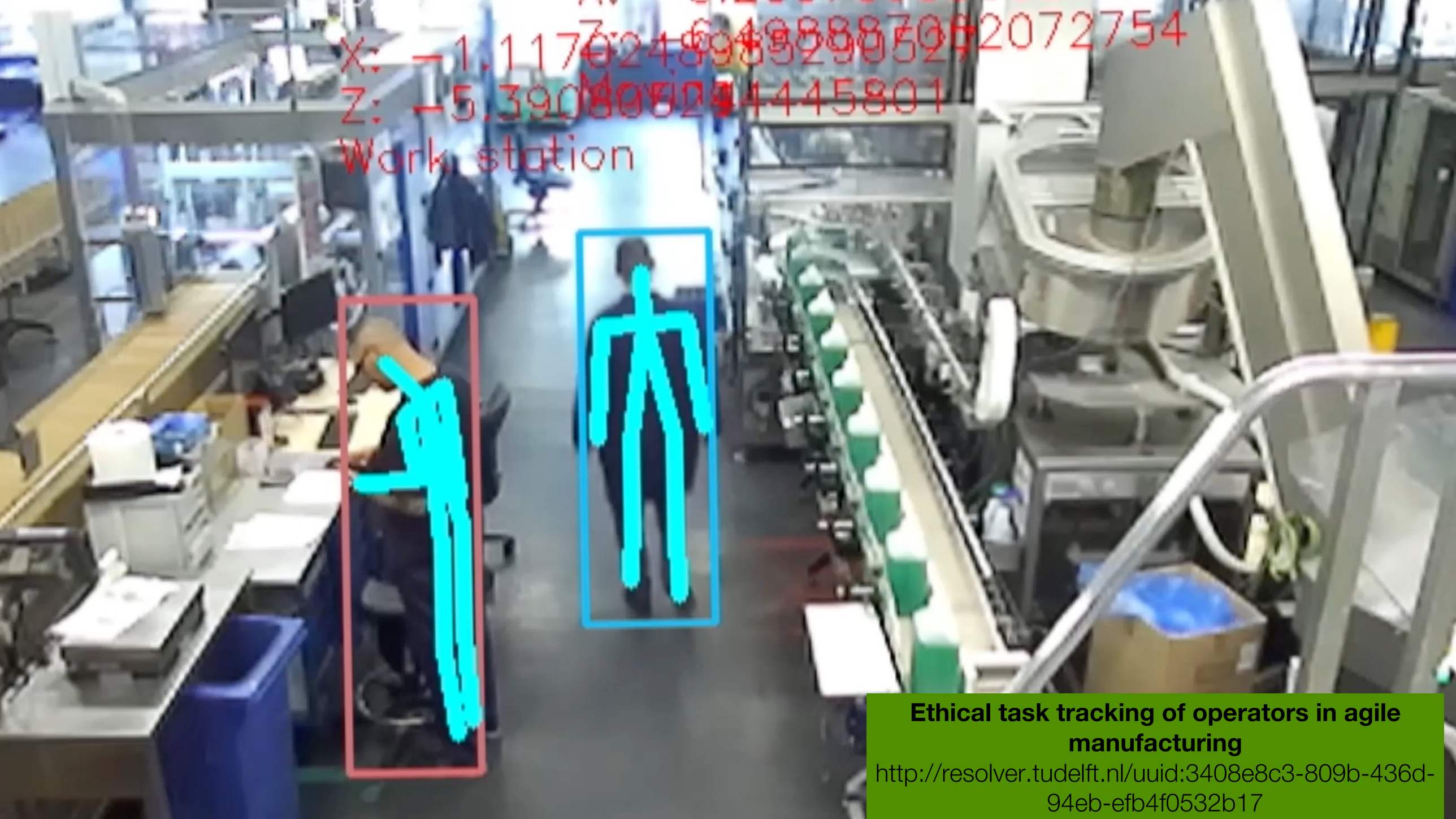
people skating on a rink in a city park with skyscrapers

Clear | Submit


What are these people doing?
(Activity / event recognition)
Skating
TRUMP
TRUMP


TRUMP
TRUMP


How are people moving?
(Pose estimation)

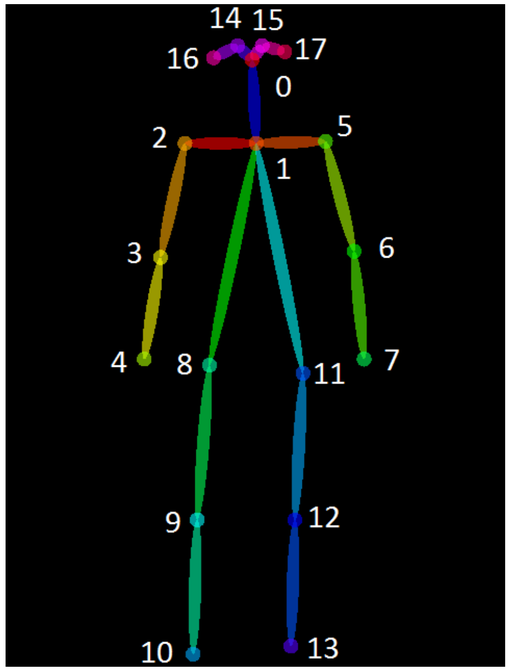


Ethical task tracking of operators in agile manufacturing

http://resolver.tudelft.nl/uuid:3408e8c3-809b-436d-94eb-efb4f0532b17

Stereolabs ZED Camera used in COALA

- 3D Object Detection
- Body tracking
- Positional tracking

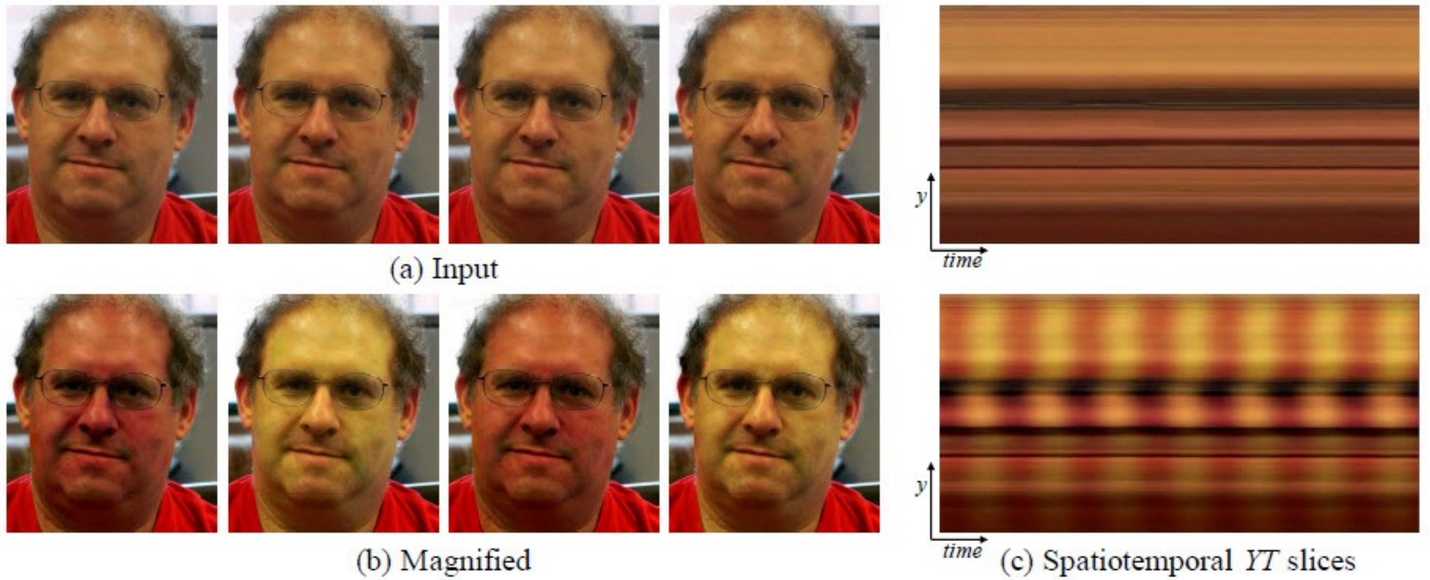


https://www.stereolabs.com/docs/object-detection/

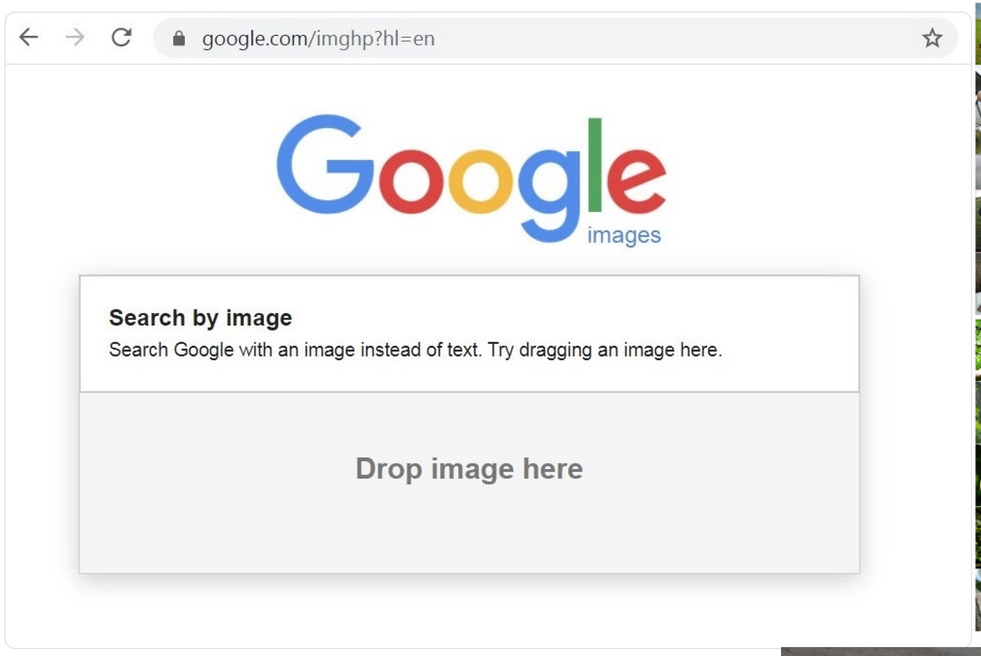
https://www.stereolabs.com/docs/body-tracking/

Eulerean Video Magnification

- Measure heart rate by magnifying and measuring the subtle changes in skin colour owed to blood vessels
- https://people.csail.mit.edu/mrub/vidmag/



Wu, H. Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., & Freeman, W. (2012). Eulerian video magnification for revealing subtle changes in the world. ACM transactions on graphics (TOG), 31(4), 1-8.



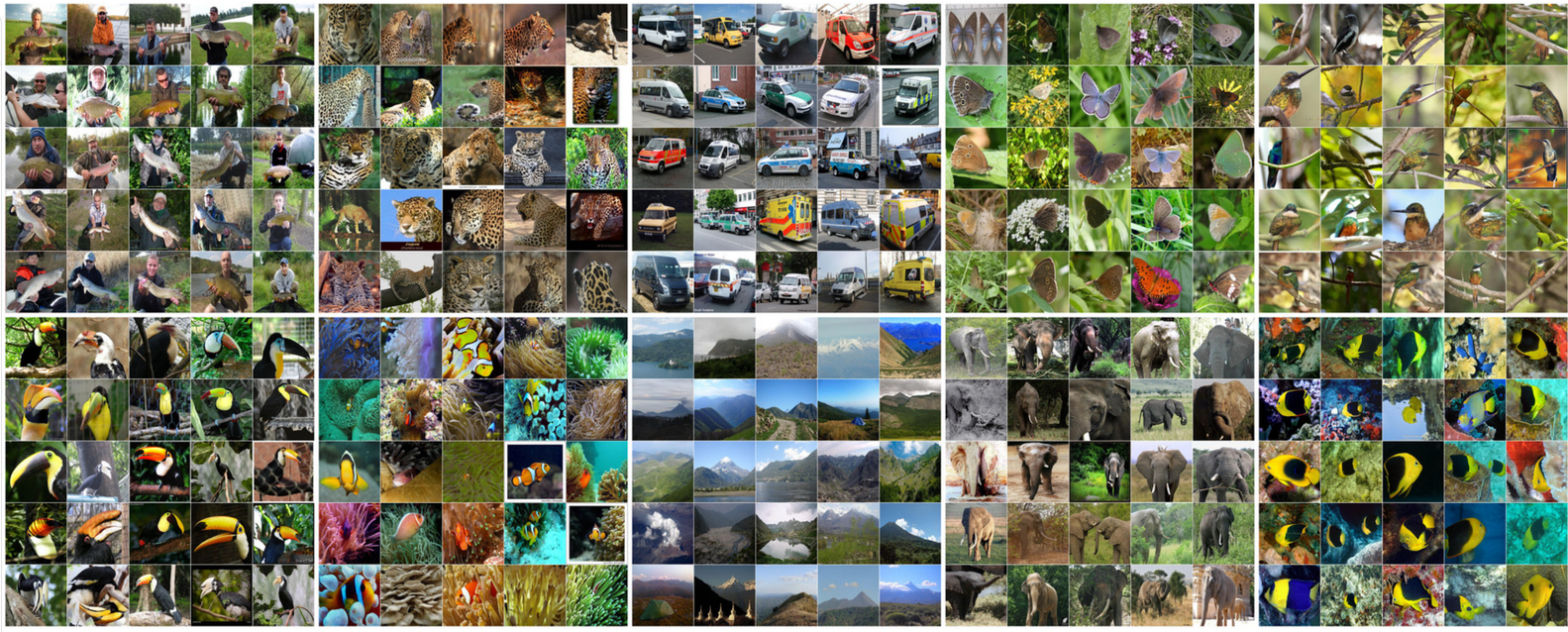
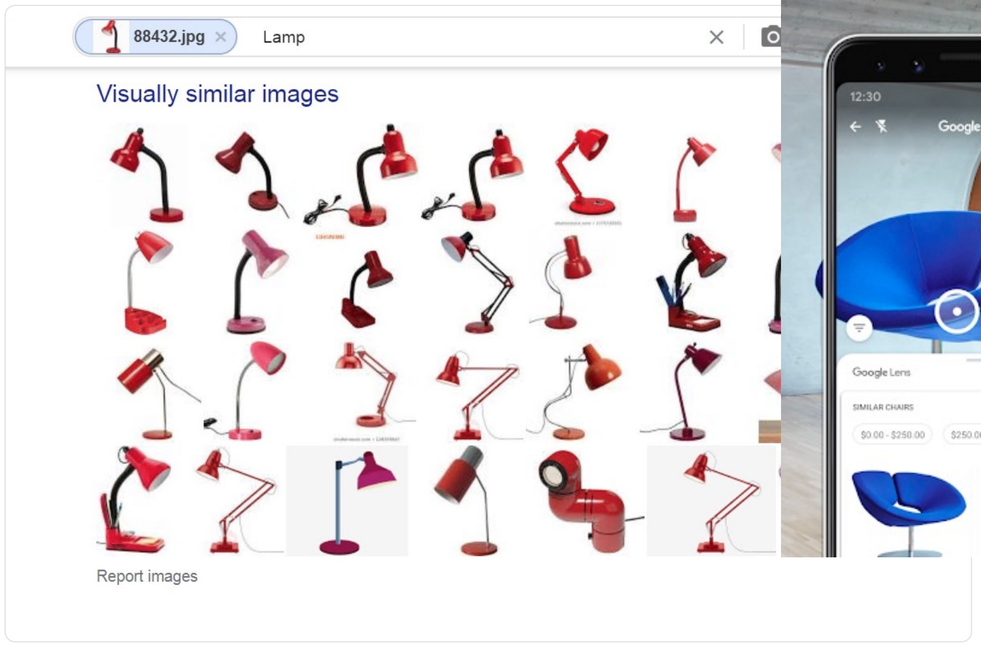
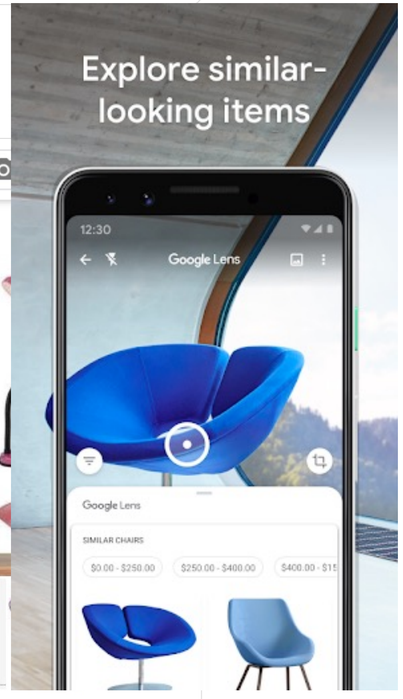
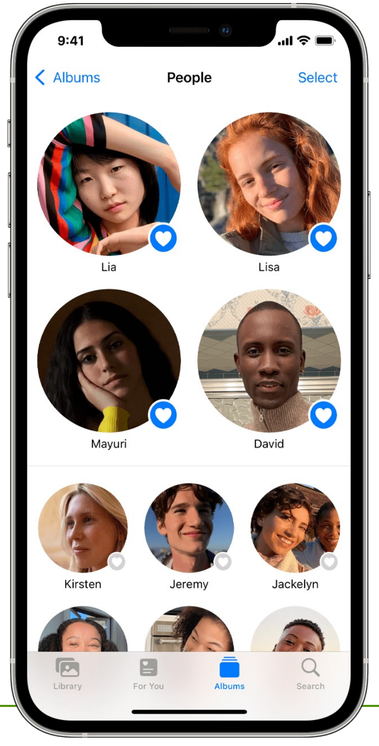


Image search results







Admin

Advanced Machine Learning For Design

Lecture 4 - Machine Learning for Image Processing (part 1)

Module 2

Evangelos Niforatos

12/10/2022

aml4d-ide@tudelft.nl
https://aml4design.github.io/

Sources

- CMU Computer Vision course - Matthew O'Toole. http://16385.courses.cs.cmu.edu/spring2022/
- Grokking Machine Learning. Luis G. Serrano. Manning, 2021
- What is back propagation really doing?"
  —YouTube walk-through
- "Gradient descent, how neural networks learn"
  —YouTube walk-through
- "Neural networks and back-propagation explained in a simple way"
  —Article on medium.com
- "The MNIST DATABASE" of handwritten digits
- "Neural Networks from Scratch"
  —YouTube tutorial
- "Neural Networks from Scratch in Python"
  —New book/e-book