Biometrics and Facial Recognition

Port of Seattle Training

July 24, 2020

University of Washington
Prof. Cecilia Aragon
http://faculty.washington.edu/aragon/
@craragon

Outline

- My background
- Definitions (biometrics and facial recognition)
- Machine learning and facial recognition
- Questions

References

- *Facial Recognition Technologies: A Primer*. Joy Buolamwini, Vicente Ordóñez, Jamie Morgenstern, and Erik Learned-Miller. May 29, 2020. https://www.ajl.org/federal-office-call
- NIST Reports on Face Recognition. Patrick Grother, Mei Ngan, and Kayee Hanaoka. Face recognition vendor test (FRVT) part 3: Demographic effects. National Institute of Standards and Technology, 2019.
  - https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf
  - https://pages.nist.gov/frvt/reports/11/frvt_11_report.pdf
- Machine learning/neural networks/deep learning tutorial (3Blue1Brown)
  - https://youtu.be/aircAruvnKk

Cecilia Aragon, PhD

- Professor, Univ. of Washington, Seattle, 2010-present
- BS, mathematics, California Institute of Technology
- PhD, computer science, UC Berkeley, 2004
- Over 15 years of software development and data science experience in industry and NASA (machine learning, image recognition, data visualization)
- 3 years as founder and CEO of small company
- Over 10 years as a professor at UW
- Over 200 publications on topics including biometrics, image recognition, data science, visual analytics, and machine learning, and 3 books published or in press
- 2008 Presidential Early Career Award for Scientists and Engineers (PECASE): the highest honor bestowed by the US government on outstanding scientists in the early stages of their careers

Definitions

- ***Biometrics*** - automated recognition of individuals based on behavioral and biological characteristics (Harmonized Biometric Vocabulary https://christoph-busch.de/standards.html)
- ***Facial recognition technologies*** – digital technologies which perform tasks on images or videos of human faces.
  - 3 types (see next slide)

Buolamwini et al, 2020.

Definitions

Facial recognition technologies – 3 types

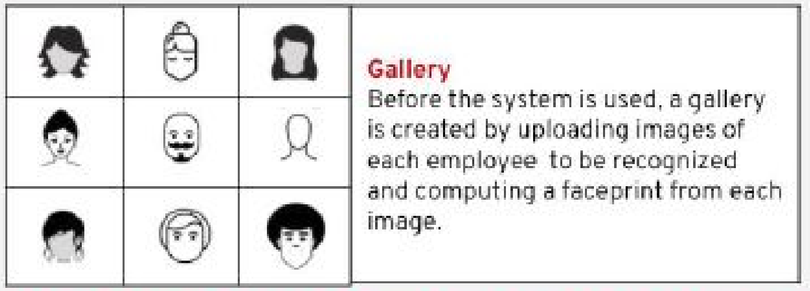
1. **Face detection**: “Is there a face in the image?”
2. **Face attribute classification**: “What kind of face is shown in the image?”
3. **Face recognition**: “Whose face is shown in the image?”
   a) **Face verification**: “Does this image show Person X?”
      - Also called *1-to-1 matching* or *1-to-1 comparison*.
      - 2 types of errors in face verification:
        - *false match (false positive)* - system incorrectly reports that images of two different people are the same
        - *false mismatch (false negative)* - system incorrectly reports that images of the same person are different.
   b) **Face identification**: “Whose face is this?”
      - Pick one person out of a *gallery* (stored appearance information of a set of people)
      - Also known as *1-to-many comparison , 1-to-many matching, 1-to-many identification, or 1-to-N identification*

This talk will focus on 1-to-1 comparison (face verification) technologies

Buolamwini et al, 2020.

Face verification (1-to-1): “Does this image show Person X?”

Figure 4: Face identification for workplace access.



Gallery
Before the system is used, a gallery is created by uploading images of each employee to be recognized and computing a faceprint from each image.

1 **Image Capture**
To gain entrance, a person poses for a picture. The captured photo is the query image.

2 **Faceprint Creation**
The system converts the query image of the person into a faceprint, a digital representation of the face.

3 **Comparison to Gallery**
The faceprint of the person is compared against the faceprints of employees in the gallery.

4 **Access Decision**
If the faceprint matches an employee, the person is allowed to enter the building. If there is no match, the person is denied entry.

The figure shows the process of using a face identification system for limiting access to a building to a set of employees.

Buolamwini et al, 2020.

Face recognition procedure

1. **Capture** – via camera or video, can be voluntary or involuntary (opt-in or opt-out)
2. **Enrollment** – recording visual information about an individual for inclusion in the gallery
3. **Faceprint generation –** digital representation of face – *faceprint (*also known as *feature vector* or *template)* – includes differentiating features, e.g. distance between eyes (but what about identical twins?)
   - Goal is to be independent of hairstyle, camera angle, image resolution, lighting, make-up, etc.
4. **Comparison –** Two faceprints are compared and a *similarity score* or *match score* is computed
   - Note: the similarity score for two different people may sometimes be higher than the similarity score for two pictures of the same person.
5. **Matching decision** based on score threshold

Trade-off between two types of errors in face verification

- Do two images show the same person (a match) or show two different people (a mismatch)?
- The system makes different decisions (shown as “Match” or “Mismatch”) by comparing the similarity score to a threshold.
- The green check marks indicate that the system’s response is correct. The red X’s show that the system has made an error.
- No single setting of the threshold eliminates all errors.
- This is a human-tunable parameter – how and when is it set?

	IMAGE PAIR	SIMILARITY SCORE	SIMILARITY SCORE THRESHOLD FOR MATCH			
			60	70	80	90
MISMATCH		65	✗ Match	✓ Mismatch	✓ Mismatch	✓ Mismatch
MATCH		73	✓ Match	✓ Match	✗ Mismatch	✗ Mismatch
MATCH		83	✓ Match	✓ Match	✓ Match	✗ Mismatch
MISMATCH		85	✗ Match	✗ Match	✗ Match	✓ Mismatch
MATCH		95	✓ Match	✓ Match	✓ Match	✓ Match
	Total False Matches (False Positives)		2	1	1	0
	Total False Mismatches (False Negatives)		0	0	1	2
	Total Error Rate		2/5	1/5	2/5	2/5

Buolamwini et al, 2020.

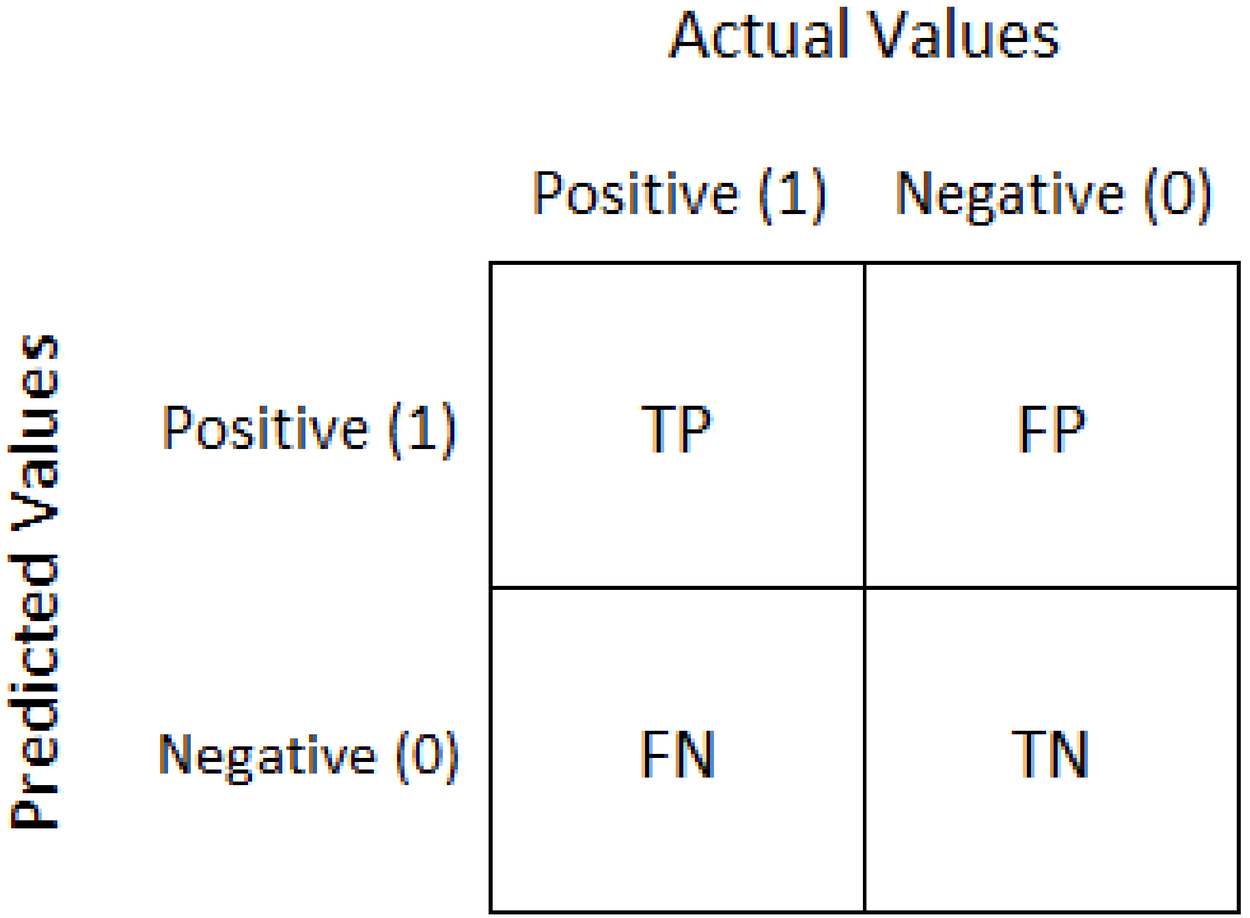
Machine learning terminology

True positive (true match) - query image matches a specific identity in a 1-to-1 comparison

True negative (or true mismatch) - imposter is unable to pass themselves as somebody else. That is, if a query image (showing Alice) is compared to Bob's passport photo, then the system's determination that they are different would be a true negative.

False positive (or false match) - the wrong person is deemed to be a match. Consequence: imposter passes through system (security concern in the 1-1 example)

False negative (or false mismatch) - rejecting the correct person. Consequence: legitimate person is denied passage (inconvenience of user in the 1-1 example)



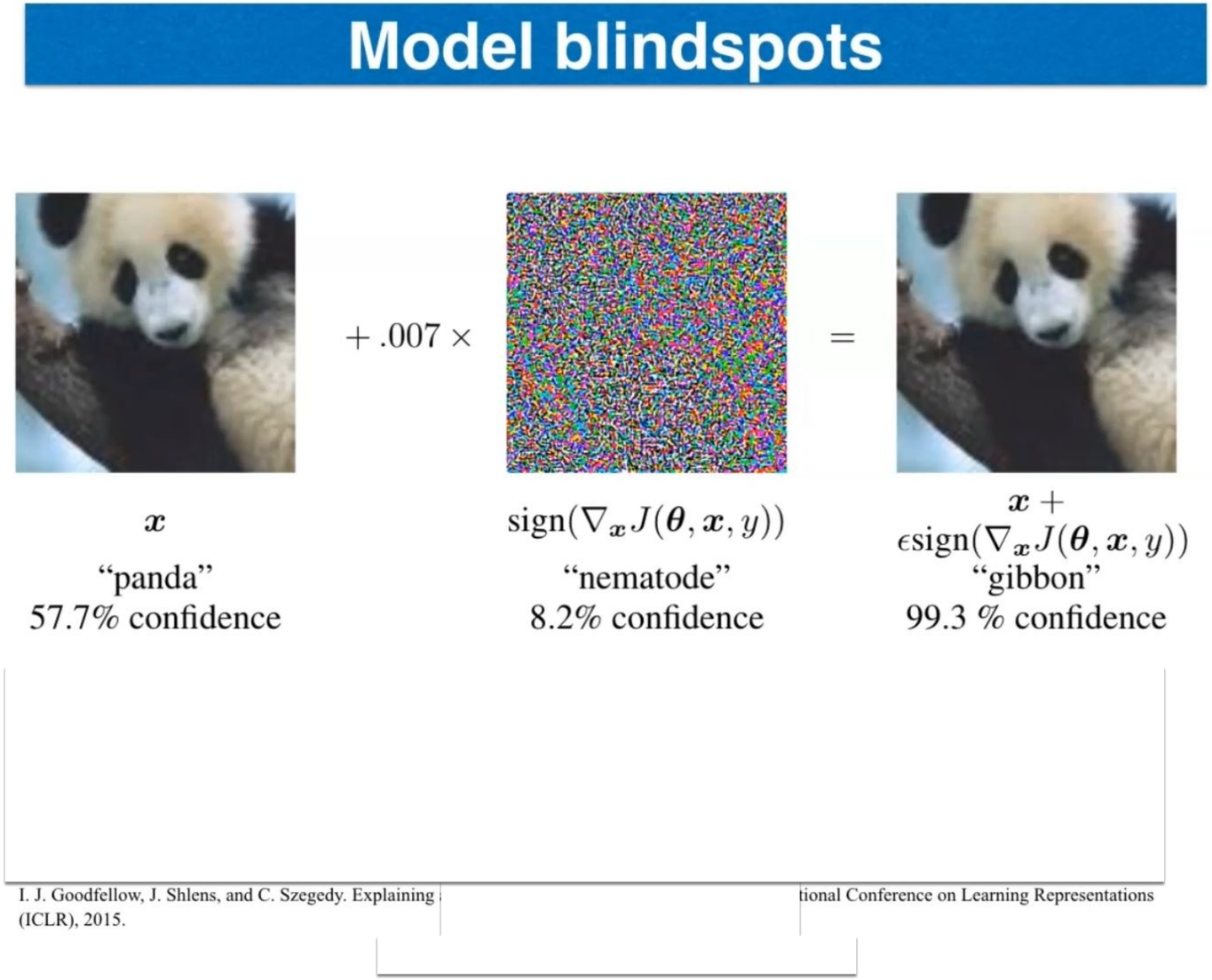
Brief History of Facial Recognition

- 1963-67 – Woody Bledsoe: reduce a face to a set of relationships between its major landmarks: eyes, ears, nose, eyebrows, lips (manual intervention)
- 1973 – Takeo Kanade: program that extracted facial features such as the nose, mouth, and eyes without human input. “The first 40 years [of facial recognition] were dominated by [Bledsoe’s] feature-based method.”
- 1990s – Sirovich and Kirby: *Eigenface* approach (linear algebra, form a set of basic features of faces from images)
- Today: Eigenfaces used as basis of many *deep learning* algorithms

Examples of Eigenfaces. Image from Scipy Lectures (https://scipy-lectures.org/packages/scikit-learn/auto_examples/plot_eigenfaces.html).

Facial recognition machine learning

- Machines don't recognize patterns the way humans do
- Famous case of algorithm recognizing image of numeral 3 as a human face
- Example: Add noise to an image, algorithm becomes "confidently wrong" (panda and gibbon example)



Suchi Saria, Microsoft Research Frontiers of Machine Learning, 2020

Machine learning and neural networks

- Excellent 20-minute tutorial: https://youtu.be/aircAruvnKk (3Blue1Brown)
- Common image recognition techniques:
  - "Vectorize" the image
    - “Unroll" the image so instead of having 2 dimensions, it only has 1

Raw Image

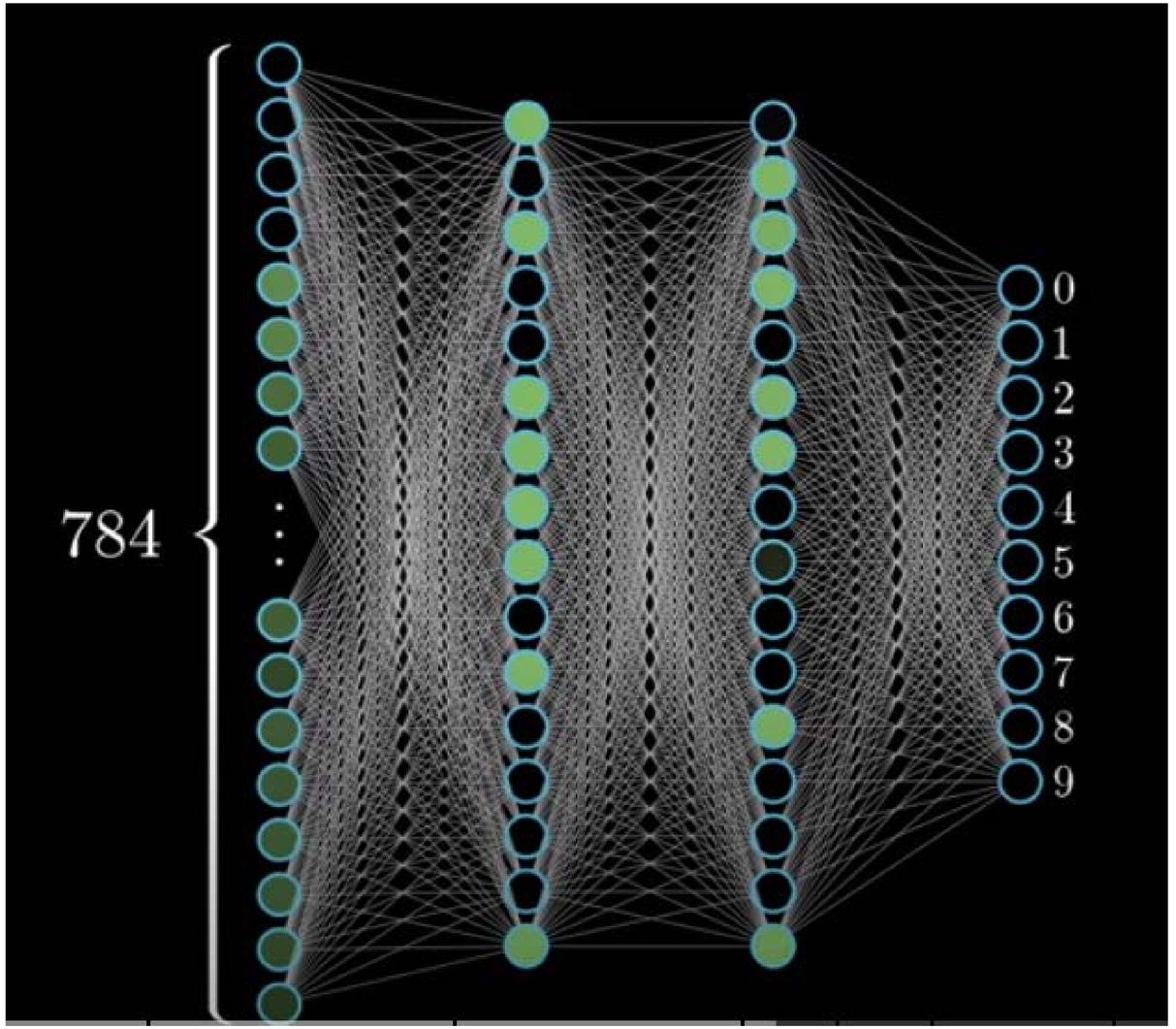
10	20	30
40	50	60
70	80	90

Unrolled Image

10
20
30
40
50
60
70
80
90

Machine learning and neural networks

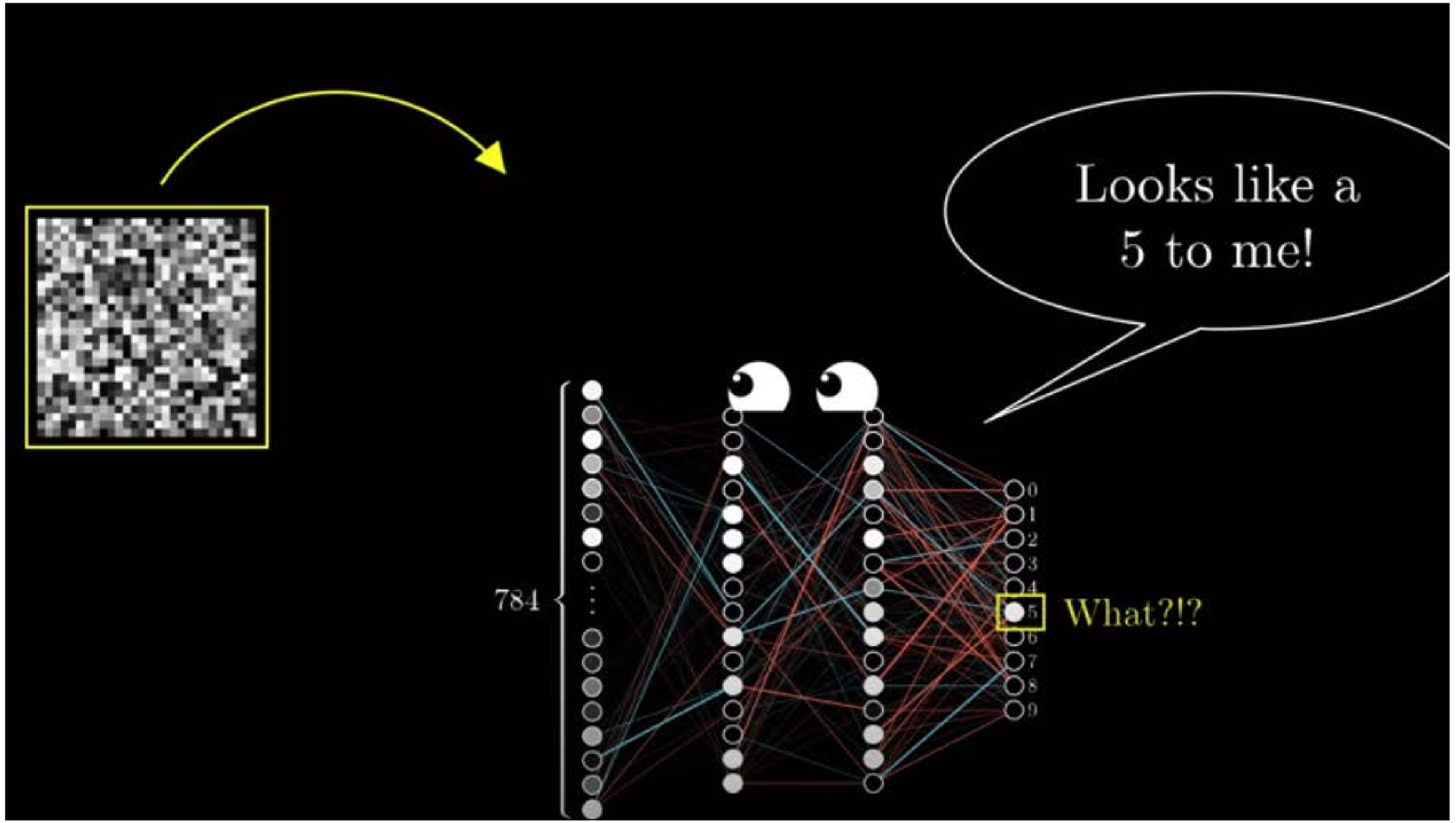
- Common image recognition techniques (cont'd):
  - *Neural networks* - powerful model class that identifies high-level concepts from low-level features like pixels
    - Developed in the 1940s, many variants and improvements since then
    - Precursor of *deep learning*, very popular and effective technique for image processing today
    - Ex: Handwritten digit recognizer, 28x28 pixels, *2 hidden layers*



https://youtu.be/aircAruvnKk

Machine learning and neural networks

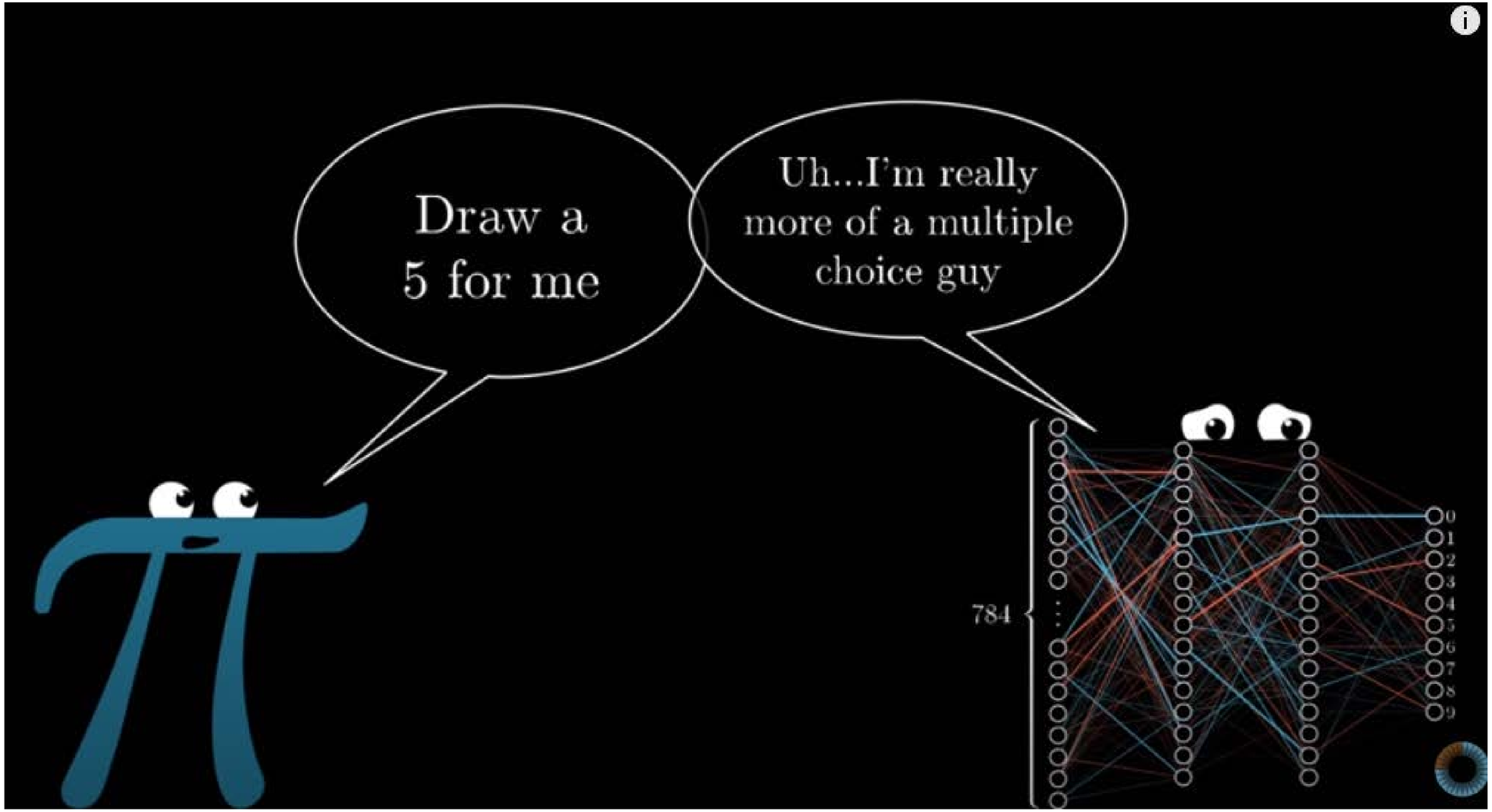
- Important to have an intuitive understanding of how machine learning works
  - Not a “magic brain that learns”!



https://youtu.be/aircAruvnKk

Machine learning and neural networks

- Another way to think of it:



https://youtu.be/aircAruvnKk

NIST Face Recognition Vendor Test 2019

- 18 million images of 8.5 million people
  - Mugshots, application photos, visa photos, border crossing photos
- 189 commercial algorithms (99 developers)
- Across demographics, false positive rates vary by factors of 10 to beyond 100 times
- False positive rates highest in West and East African and East Asian people, and lowest in Eastern European individuals
  - Algorithms developed in China: effect is reversed, with low false positive rates on East Asian faces
  - Domestic law enforcement images: highest false positives in Indigenous populations, elevated rates in African American and Asian populations
  - False positives higher in women than men
- False negatives in mugshots higher in Asian and Indigenous individuals, but in lower-quality border crossing images, false negatives higher in people born in Africa and the Caribbean (effect stronger in older individuals)

Patrick Grother, Mei Ngan, and Kayee Hanaoka. Face recognition vendor test (FRVT) part 3: Demographic effects. National Institute of Standards and Technology , 2019.

Addressing your questions

1. How accurate is a face recognition system?

- Machine learning model performance metrics
- Benchmarks

Machine learning model performance metrics

- *Precision* = number of correct positive results divided by the number of positive results predicted by the classifier
- *Recall* = number of correct positive results divided by the number of all relevant samples
- High precision but lower recall means an accurate model which at the same time misses a large number of instances that are difficult to classify.
- *F1* = harmonic mean of precision and recall
  - attempts to provide a metric that accounts for both how precise and robust the model is

https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234

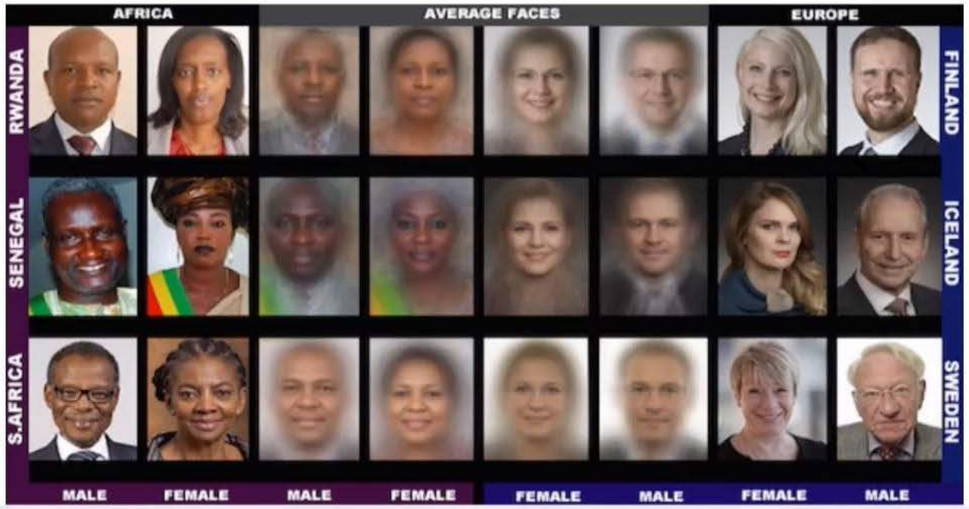
Performance metrics and benchmarks

- Complex issue, many variables, need to consider context
- Who sets the parameters in the real world? How often?
- Can’t assume that the statistics measured on a standard benchmark will be representative of the system’s performance in real-world scenarios
- NIST 2019:
  - Benchmark test on diverse population had 1 in 10,000 error rate (false match)
  - Same system on more homogeneous population: false match rates 20 times higher
  - Why?
    - When people share more traits, they can be harder to distinguish

Patrick Grother, Mei Ngan, and Kayee Hanaoka. Face recognition vendor test (FRVT) part 3: Demographic effects. National Institute of Standards and Technology , 2019.

2. What factors impact the accuracy of facial recognition technology?

- Factors such as different settings, different sub-populations, lighting, image quality, facial orientation, occlusion, and camera motion can have dramatic effects on the results
- Setting of threshold is crucial (and can't avoid errors)
- Buolamwini et al. study on gender classifiers found differences in error rates between darker/lighter female/male populations
- National Institute of Standards and Technology (NIST) December 2019 report showed differences across demographic groups



J. Buolamw

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

3. Is algorithmic bias inevitable?

- A better question might be: is it possible to mitigate algorithmic bias, and if so, what procedure can we use to do so?
- This is an active area of machine learning research.

Framework to Engineer for Reliability

Failure Prevention | **Test-time Monitoring** | **Maintenance**

Clear system specification and understanding of expected behavior.

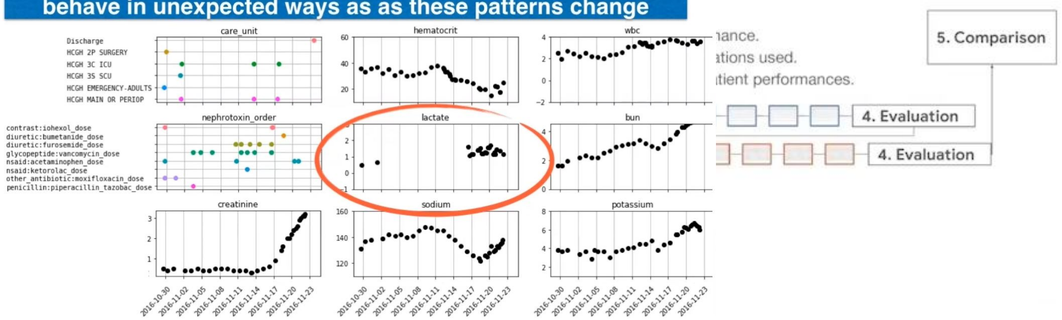
1. **Failure Prevention:** Approaches that prevent or reduce the likelihood of failures or unexpected behaviors
2. Identify failures and their causes when they occur in real-time [**Failure identification**] and [**Reliability Monitoring**]
3. Fix the failures when they occur [**Maintenance**]

Saria, Subbaswamy, **Tutorial: Safe and Reliable Machine Learning**. ACM Fairness, Accountability and Transparency, 2019.

Human and Machine Learning In The Loop

- New opportunities to improve **collaboration** between models and clinicians.

Practice Evolves Over Time and naively trained ML models can behave in unexpected ways as as these patterns change



- Goal: Use labs to predict risk of an adverse event
- Trained on data from 2011-2013 and tested on 2014, it performed very well. When tested on 2015, performance deteriorated dramatically.
- Instance of learning a dependency that does not generalize across changes in provider ordering patterns.

Schulam, P and Saria, S. "Reliable Decision Su… Models", Neural Information Processing Systems, 2017.

KNOW, THIS IS IN A HOSPITAL SETTING

Questions?

Cecilia Aragon

University of Washington

http://faculty.washington.edu/aragon/

@craragon