

GR5074

Projects in Advanced Machine Learning

Thursdays, 6:10-8pm

253 Engineering Terrace

Instructor

Alexander Peterhansl (he/him)

Fun Facts About Me

- Born in Iran to an Iranian mother and a German father.
- Love affair with New York
- Worked on Wall Street in FX market-making and analytics, as well as the art and finance market.
- Used to swim competitively and run marathons. Now I am grateful to get on a Citi bike and ride around with my two kids.

.

Course Arrangement

- 60-75 minute lecture
- followed by 5 minute break
- followed by 30-45 minute recitation by teaching assistant

Syllabus

- link

Machine Learning Versus Conventional Programming

- Early approach to face identification
- The state of the art in 1965 – secretly funded by the CIA – initially used a conventional programming approach

Face Identification



RAND Grafacon in 1965



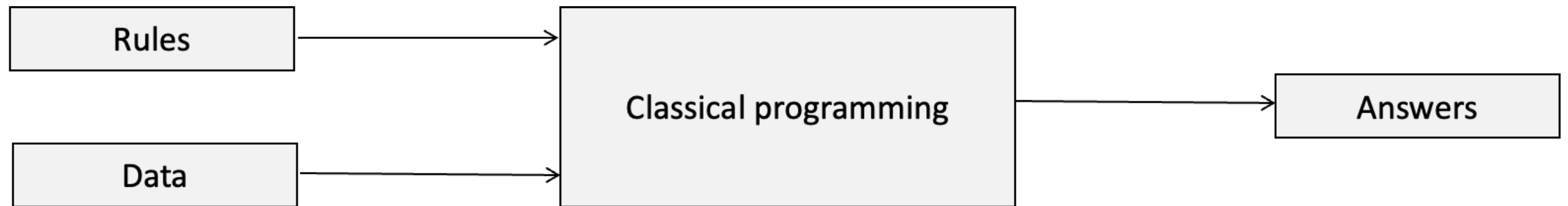
Face Identification in 1965

- RAND Project
- Officially referred to as a "man-machine" project – why?
- Volunteers marked 20 distances on a RAND GRAFACON (1965 version of a tablet computer):
 - Width of eyes
 - Pupil-to-pupil distance
 - Width of mouth, etc.
- Set of distances for each image were matched to 'closest' set of corresponding distances in database record

What is the machine learning approach?

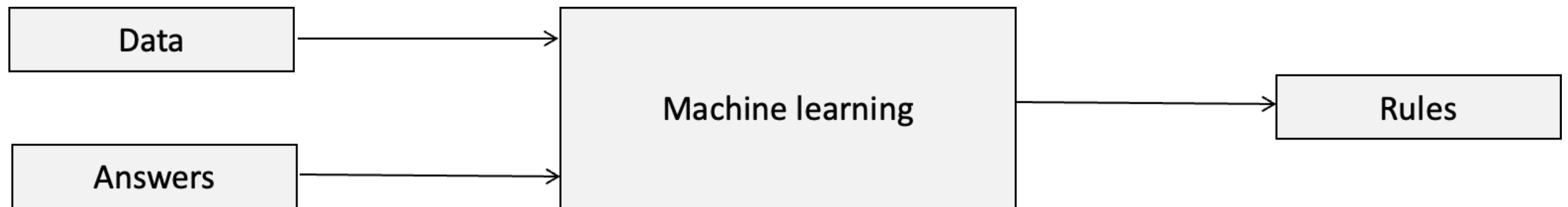
- What is one key aspect that both approaches have in common?
 - Both rely on a mathematical representation of the face.
- What are the differences?
 - Traditional approach relies on a model that is fixed **ex-ante**!
 - Traditional model does not improve with more data. It gives a better fit, but the model is the same.
 - With traditional model, hard to differentiate between a bad model and 'hard-to-identify' images (faces that are smiling, frowning, grimacing, turned at an angle)
 - Machine learning addresses all these shortcomings. How?
- What is the downside?
 - Much harder to pin down why something does not work well in the new method.

Classical Programming Schema



- How does the 1965-approach fit into this schema?

Machine Learning Schema



- Which key step falls away in this approach?
- What is it replaced by?

Neural Network Workhorse

- X_j is a variable / input / feature
- w_j parameter / weight to be estimated

$$A = h(X) = g \left(w_0 + \sum_{j=1}^p w_j X_j \right)$$

- $A = h(X)$ is the (non-linear) activation function