Pattern Classification

01. Introduction

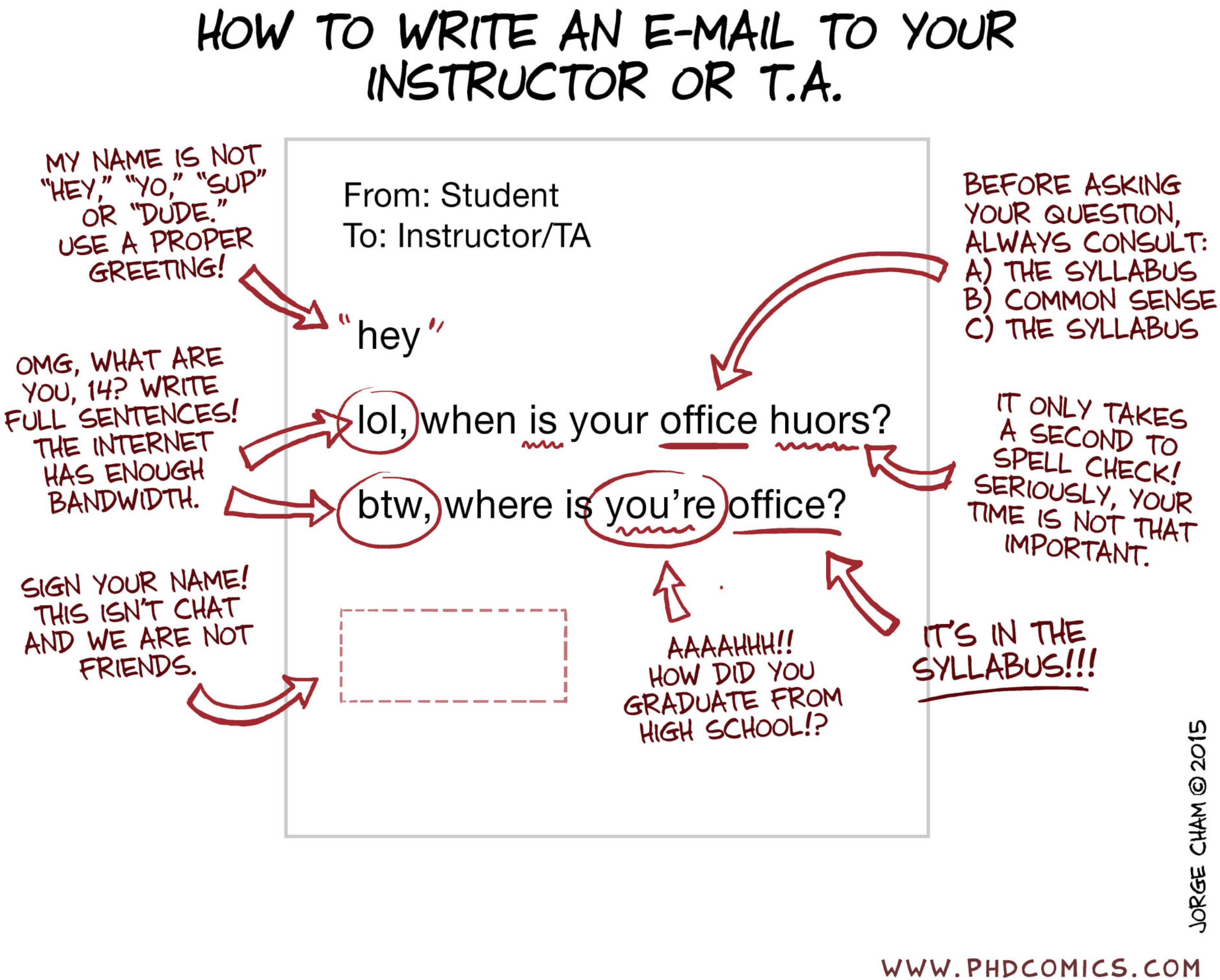
AbdElMoniem Bayoumi, PhD

Spring 2023

Administrivia

- Contacts:
  - abayoumi@cu.edu.eg
- Grading Policy:
  - Midterm: 10%
  - Labs and Assignments: 5%
  - Project: 25% (tba latter)
  - Final Exam: 60% (written & closed book exam)
- Shared Folder: https://shorturl.at/gS357

Email Guidelines!



References

- Deep Learning, MIT Press, by Ian Goodfellow, Yoshua Bengio, Aaron Courville
  - https://www.deeplearningbook.org/

- Introduction to statistical pattern recognition, 2nd edition, by Keinosuke Fukunaga

What is AI?

- Views of AI fall into four categories:

Thinking humanly	Thinking rationally
Acting humanly	Acting rationally

  - Most of computer scientists consider acting rationally

Thinking Humanly: Cognitive Science

- **Cognitive revolution** in 1960s:
  - Information-processing psychology replaced prevailing behaviorism
- Requires scientific theories of internal activities of the brain:
  - What level of abstraction? “Knowledge” or “circuits”?
- Validation:
  - Predicting and testing behavior of human subjects (top-down) → **Cognitive Science**
  - Direct identification from neurological data (bottom-up) → **Cognitive Neuroscience**
- Both approaches (Cognitive Science and Cognitive Neuroscience) are now distinct from AI

Thinking Rationally: Laws of Thought

- Aristotle: What are correct arguments / thought processes?

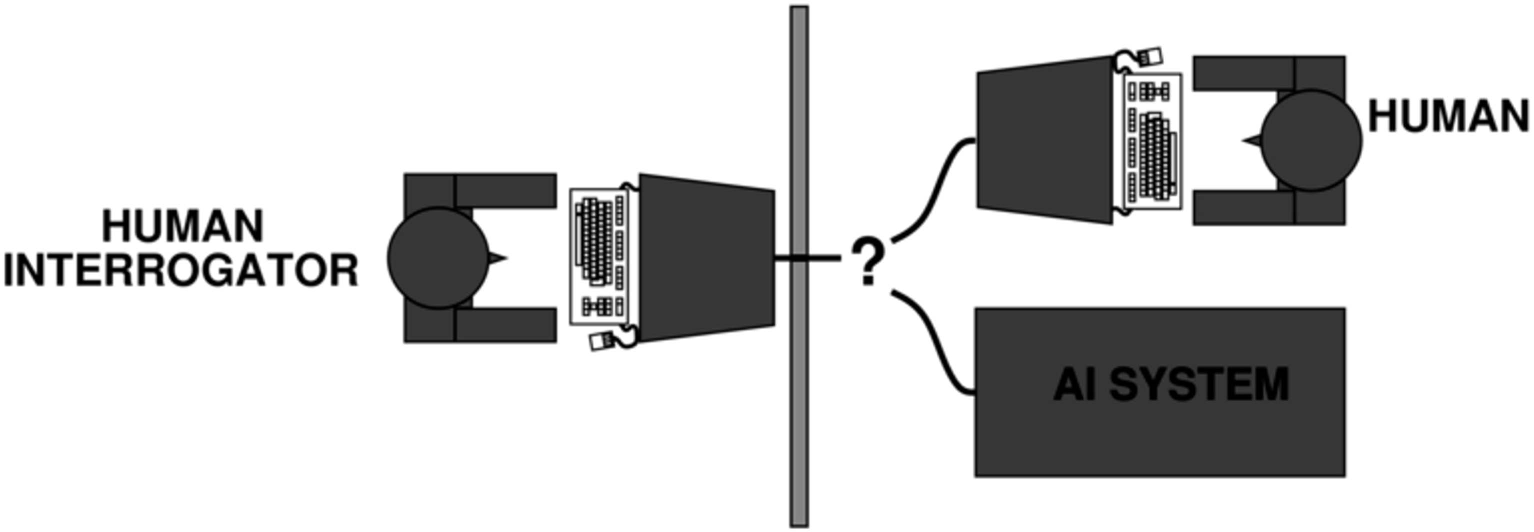
- Several Greek schools developed various forms of **logic**:
  - Notation and rules of derivation for thoughts

- Direct line through mathematics and philosophy to modern AI

- Not all intelligent behavior is mediated by logical deliberation

Acting Humanly: The Turing test

- Alan Turing in his "Computing machinery and intelligence" paper in 1950:
  - "Can machines think?" → "**Can machines behave intelligently?**"
  - Operational test for intelligent behavior: *the Imitation Game*



Acting Humanly: The Turing test

- Predicted that by 2000, a machine might have a 30% chance of fooling a lay person for 5 minutes
- Anticipated all major arguments against AI in following 50 years
- Suggested major components of AI: knowledge, reasoning, language understanding, learning
- Problem: Turing test is **NOT** *reproducible*, *constructive*, amenable to *mathematical analysis*

Acting Rationally

- Rational behavior: doing the right thing
- The right thing: that which is **expected to maximize goal achievement**, given the available information
- Doesn't necessarily involve thinking (e.g., blinking reflex) but thinking should be in the service of rational action
- Aristotle: “Every art and every inquiry, and similarly every action and pursuit, is thought to aim at some good”

Applications of AI

- Self-driving cars:
  - AI is still not ready yet for it (Tesla crashes)

Tesla Self-Driving Car:
https://www.tesla.com/en_CA/autopilot?redirect=no

Applications of AI

- Facebook:
  - face recognition (violate user's privacy?!)
  - personalized Ads
  - Cambridge Analytica (US elections 2016)

- Google:
  - search engine results refining
  - Gmail responds & smart compose: https://www.youtube.com/watch?v=nZ-C8I-8BZw
  - Google Duplex: https://www.youtube.com/watch?v=D5VN56jQMWM
  - https://ai.google/

- Twitter:
  - recommendations
  - filter racist & inappropriate content

Applications of AI

- Amazon:
  - visual search option
  - warehouse robots: https://www.youtube.com/watch?v=UtBa9yVZBJM

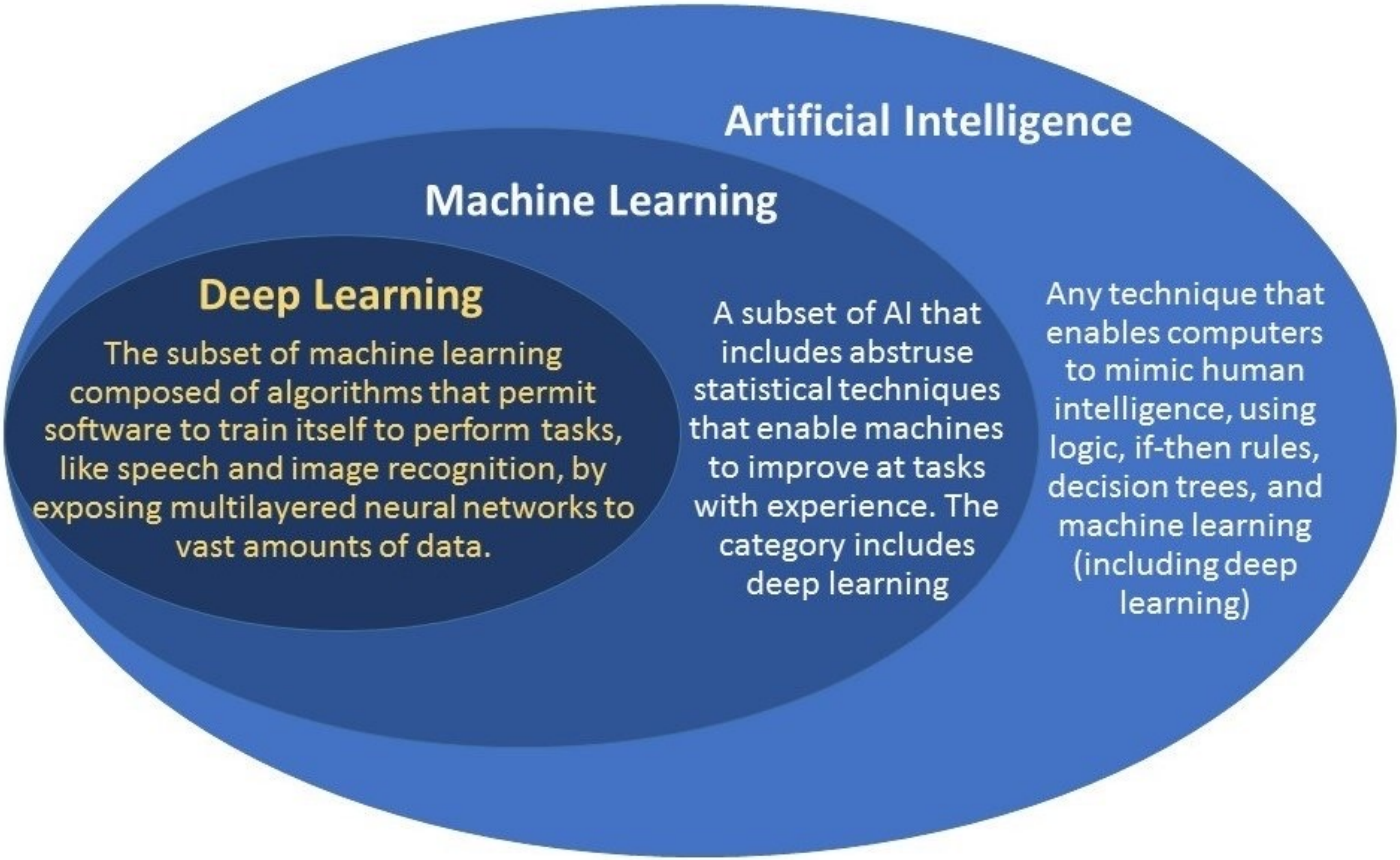
- Virtual Personal Assistants:
  - Siri, Alexa, Google Now

- Video Surveillance Systems:
  - motion detection
  - sound detection (boom or window crashing)
  - https://youtu.be/7kIDMEOc1nA

Subfields of AI

- AI now consists many sub-fields, using a variety of techniques, such as:
  - Neural Networks – e.g. brain modeling, time series prediction, classification
  - Evolutionary Computation – e.g. genetic algorithms, genetic programming
  - Computer Vision – e.g. object recognition, image understanding
  - Robotics – e.g. intelligent control, autonomous exploration
  - Expert Systems – e.g. decision support systems, teaching systems
  - Speech Processing– e.g. speech recognition and production
  - Natural Language Processing – e.g. machine translation
  - Planning – e.g. scheduling, game playing

Subfields of AI

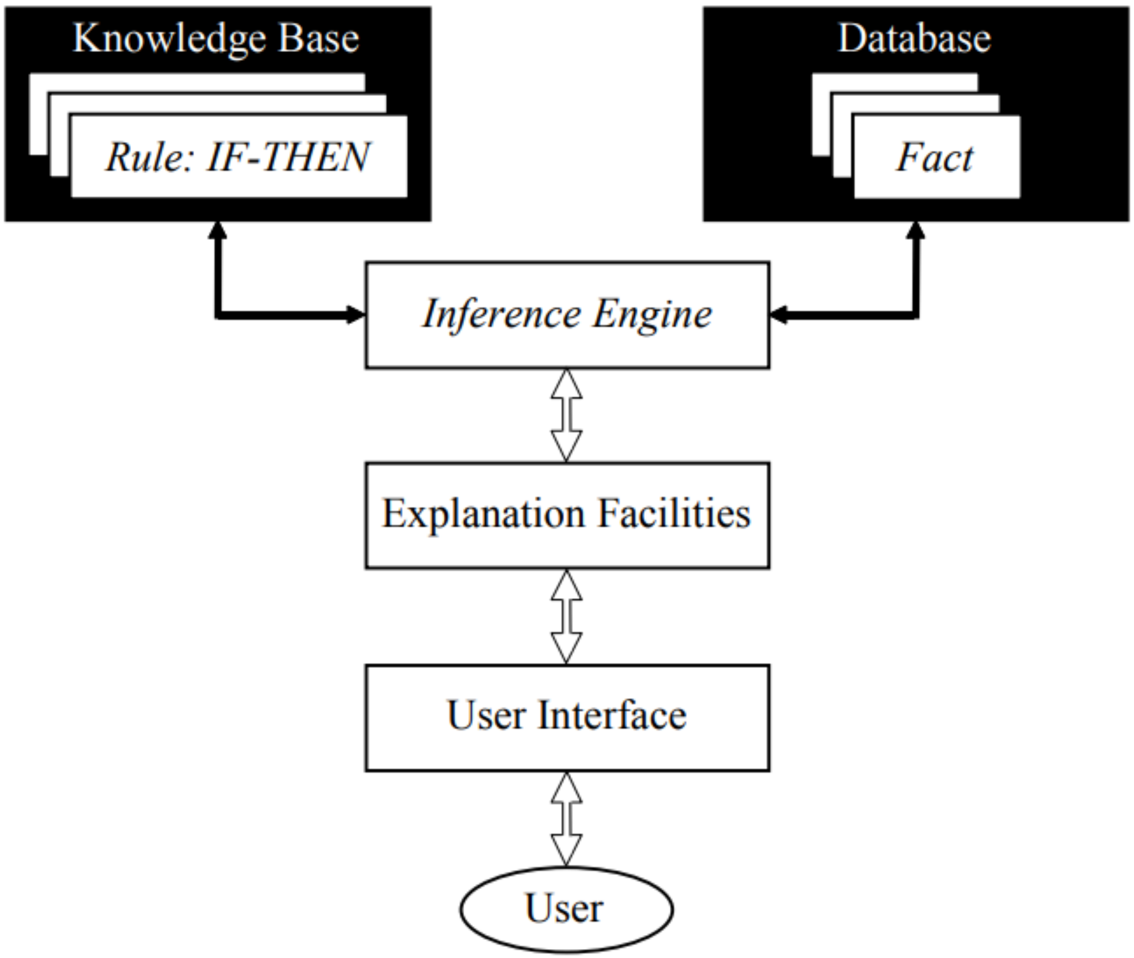


Classic Artificial Intelligence

- AI includes any technique that mimics human behavior
- Classic AI aka. AI excluding machine learning (and deep learning as well)
  - Includes rule-based expert systems
  - Knowledge reasoning & inference
  - No learning needed (or no data is available!)
  - Simpler applications

Classic Artificial Intelligence

- Expert System



Source: Negnevitsky, Pearson Education, 2002

Artificial Intelligence

- Example application: Roomba robot vacuum cleaners
  - Infrared sensors to detect obstacles
  - Touch-sensitive bumpers to mitigate collision
  - Piezoelectric sensor for dirt detection
  - Camera to build a map using V-SLAM

Classic Artificial Intelligence

- Example application: AI in video games
  - Decision trees
  - Path finding

Machine Learning

- Subset of AI
- Machine learning algorithms provide computers with learning skills
  - Computers programs learn from examples
  - Improve performance based on experience
- Includes statistical techniques

Machine Learning

- Indirect instructions
  - describe a friend’s face to some other person vs. showing him example pictures (i.e., face recognition)

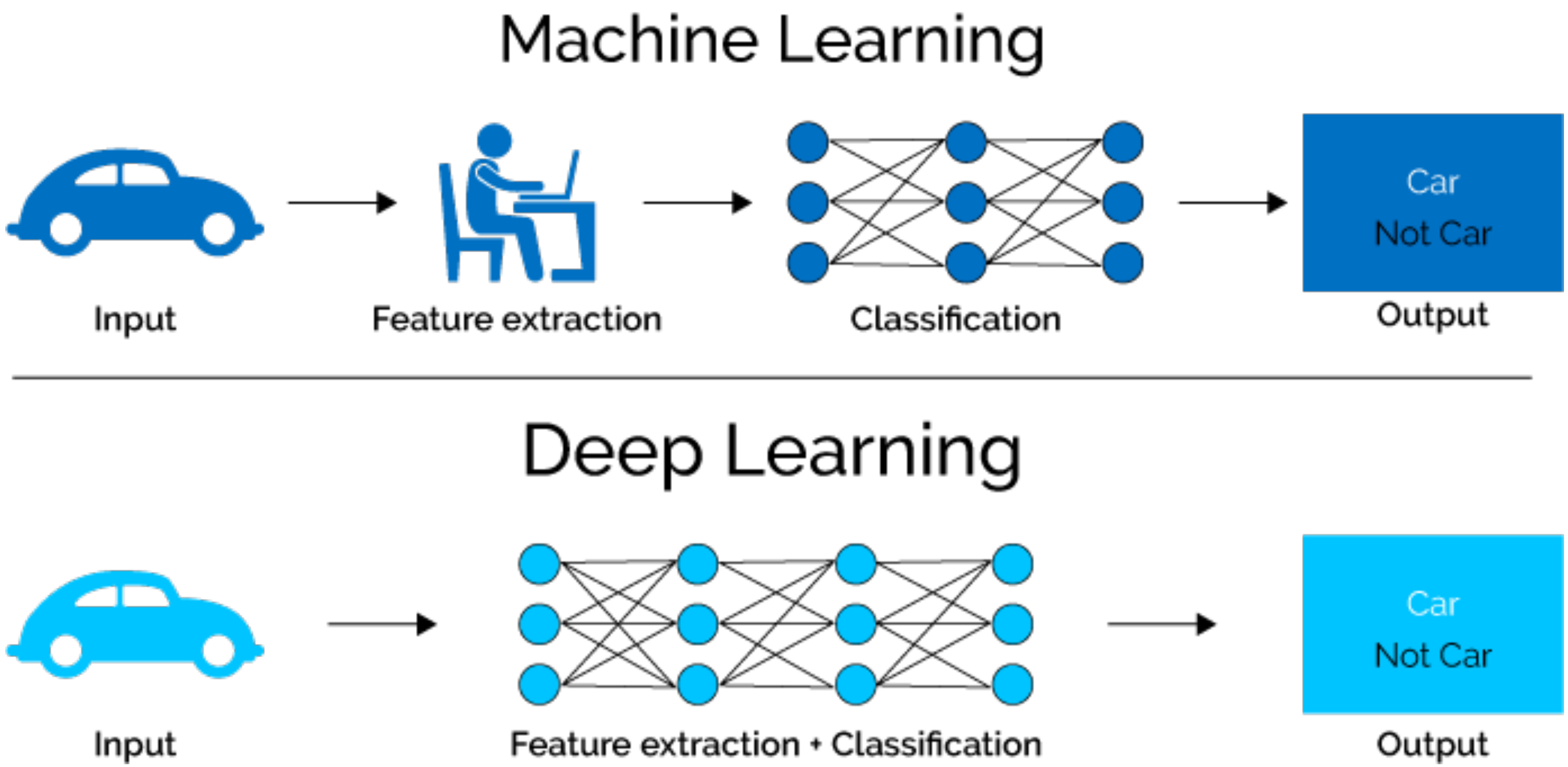
- Finding patterns in data and make use of that, i.e., function approximation

- Adapt to complex and changing environments

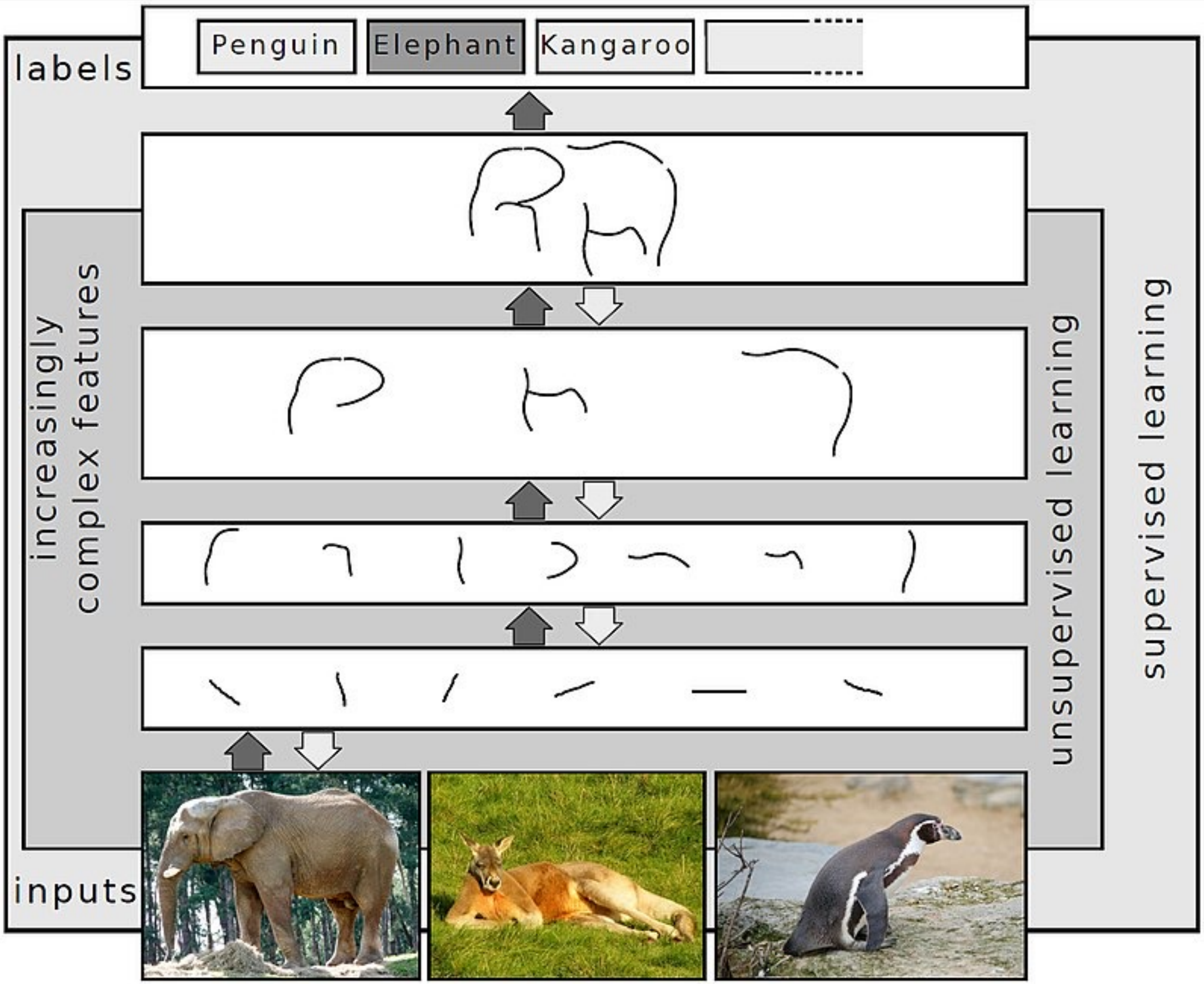
Deep Learning

- Subset of machine learning
- Multi-layered neural networks
- Raw data, i.e., end-to-end solution
- Requires big data & high computational power

Machine Learning vs Deep Learning

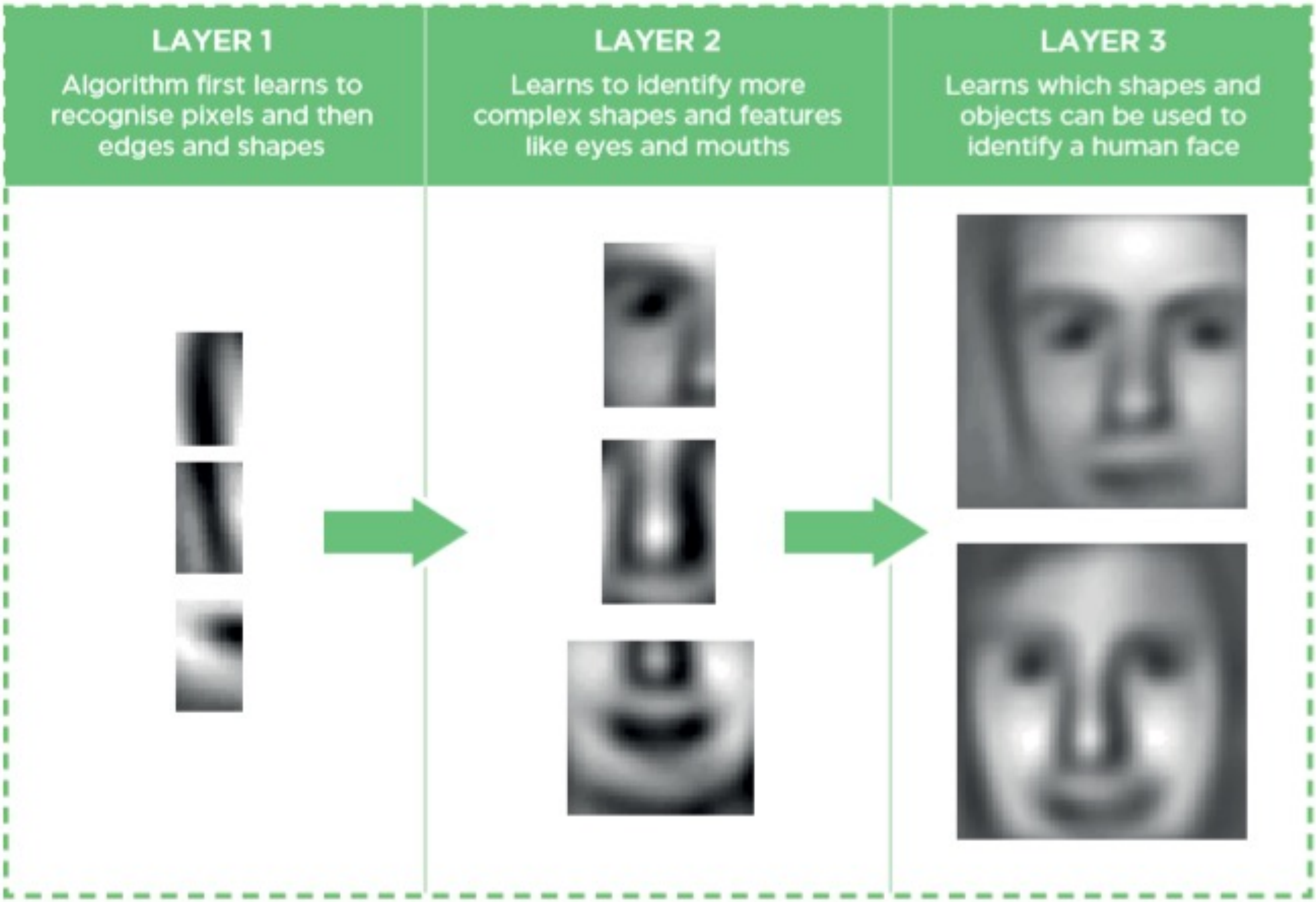


Machine Learning vs Deep Learning



Source: Hannes Schulz and Sven Behnke, 2012

Machine Learning vs Deep Learning



When to use Deep Learning?

- Big amount of **data**

  expensive!

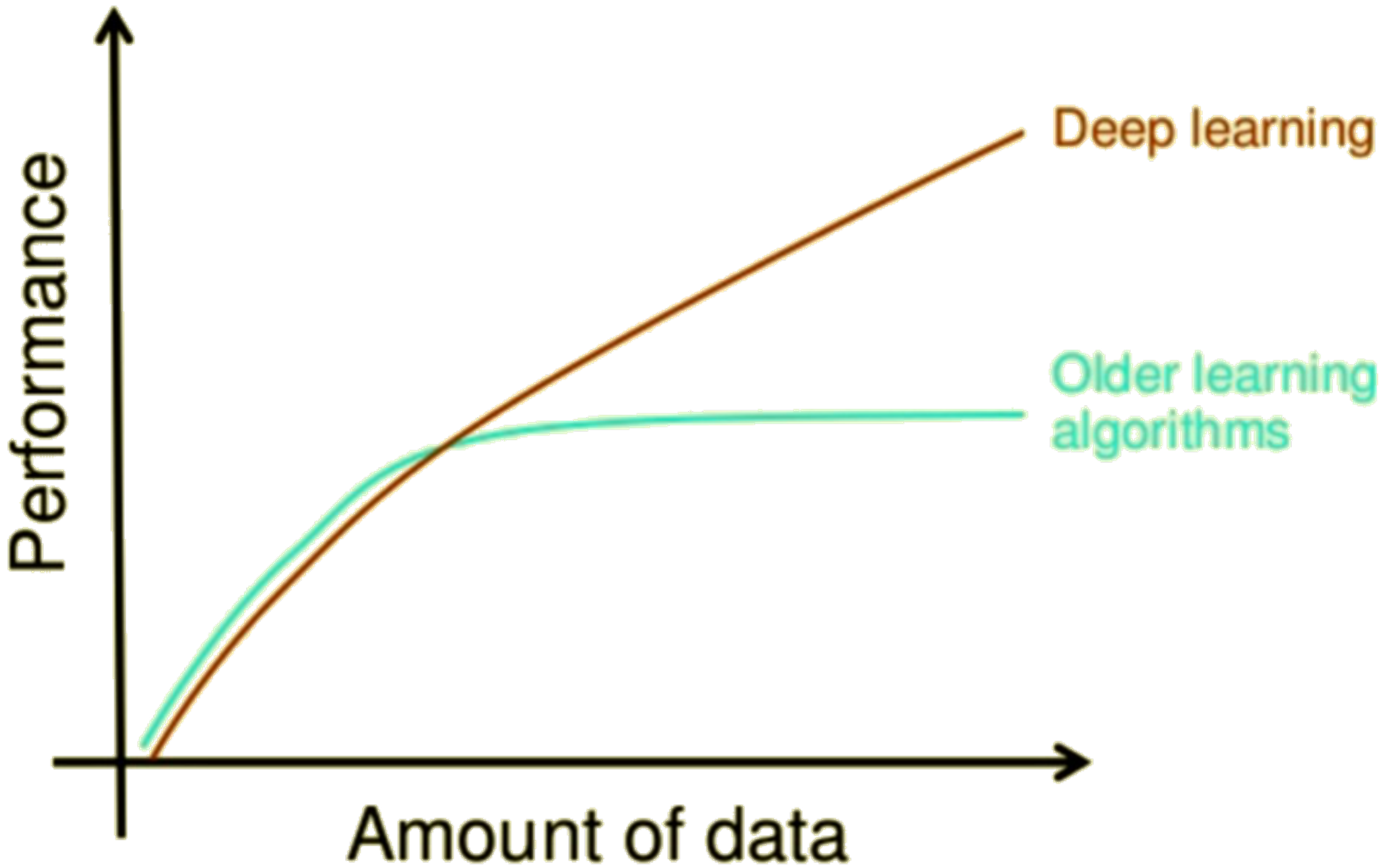
- Availability of high computational power

  expensive!

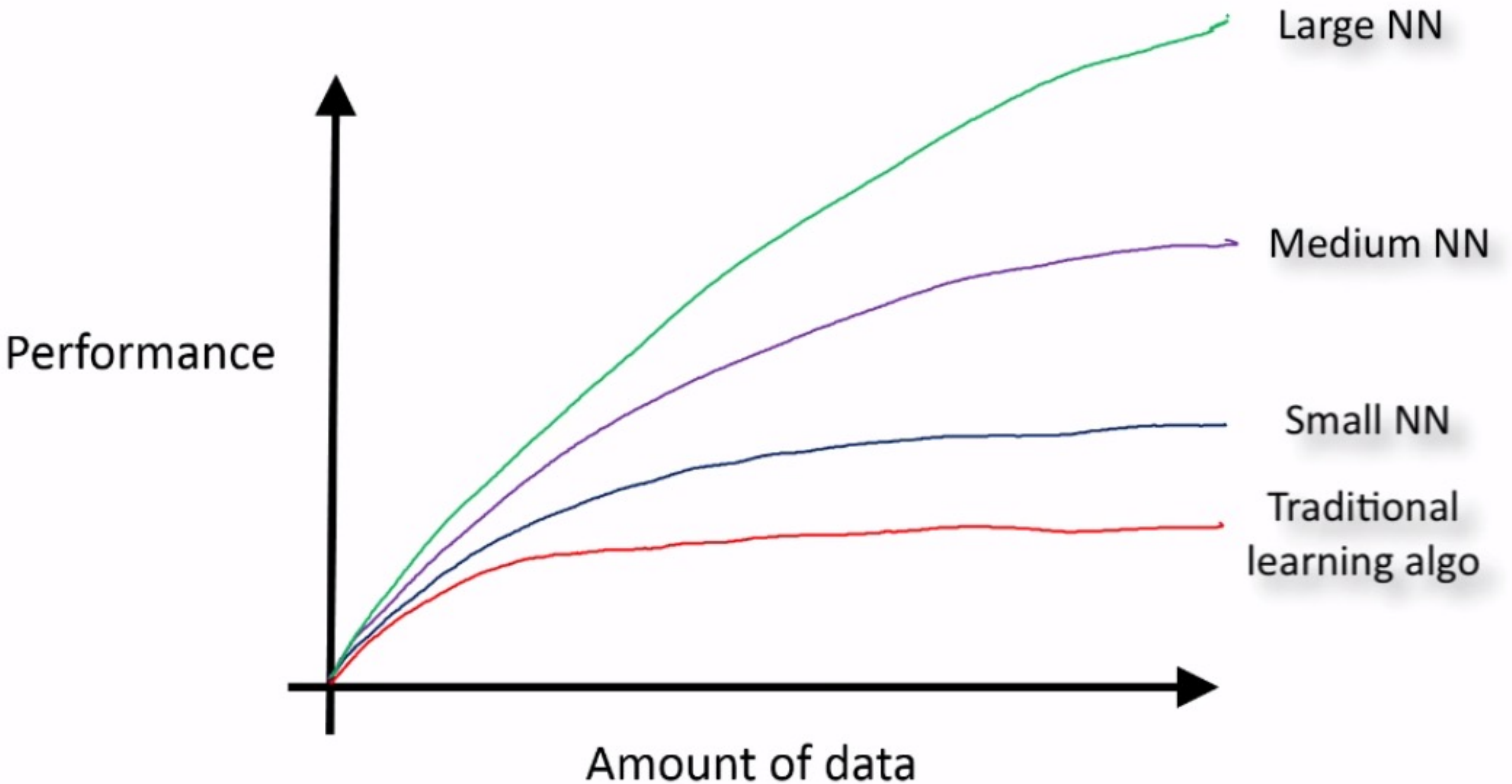
- Lack of domain understanding

- Complex problems (vision, NLP, speech recognition)

Scalability with Data Amount



Scalability with Data Amount



Andrew Ng

Potentials of AI

"If a typical person can do a mental task with less than one second of thought, we can probably automate it using AI either now or in the near future."

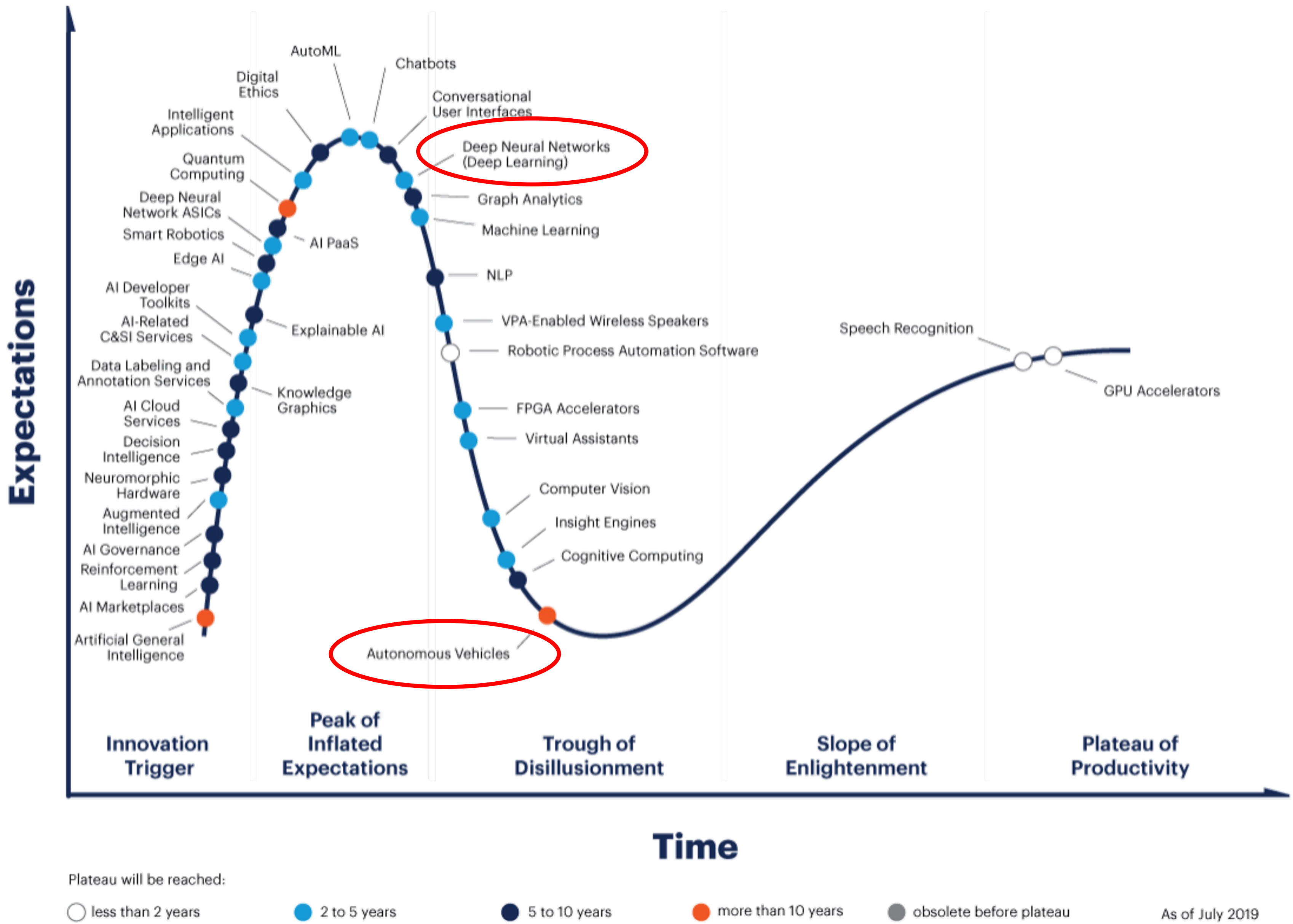
— Andrew Ng

Currently, there are some limitations!

Limitations of Deep Learning

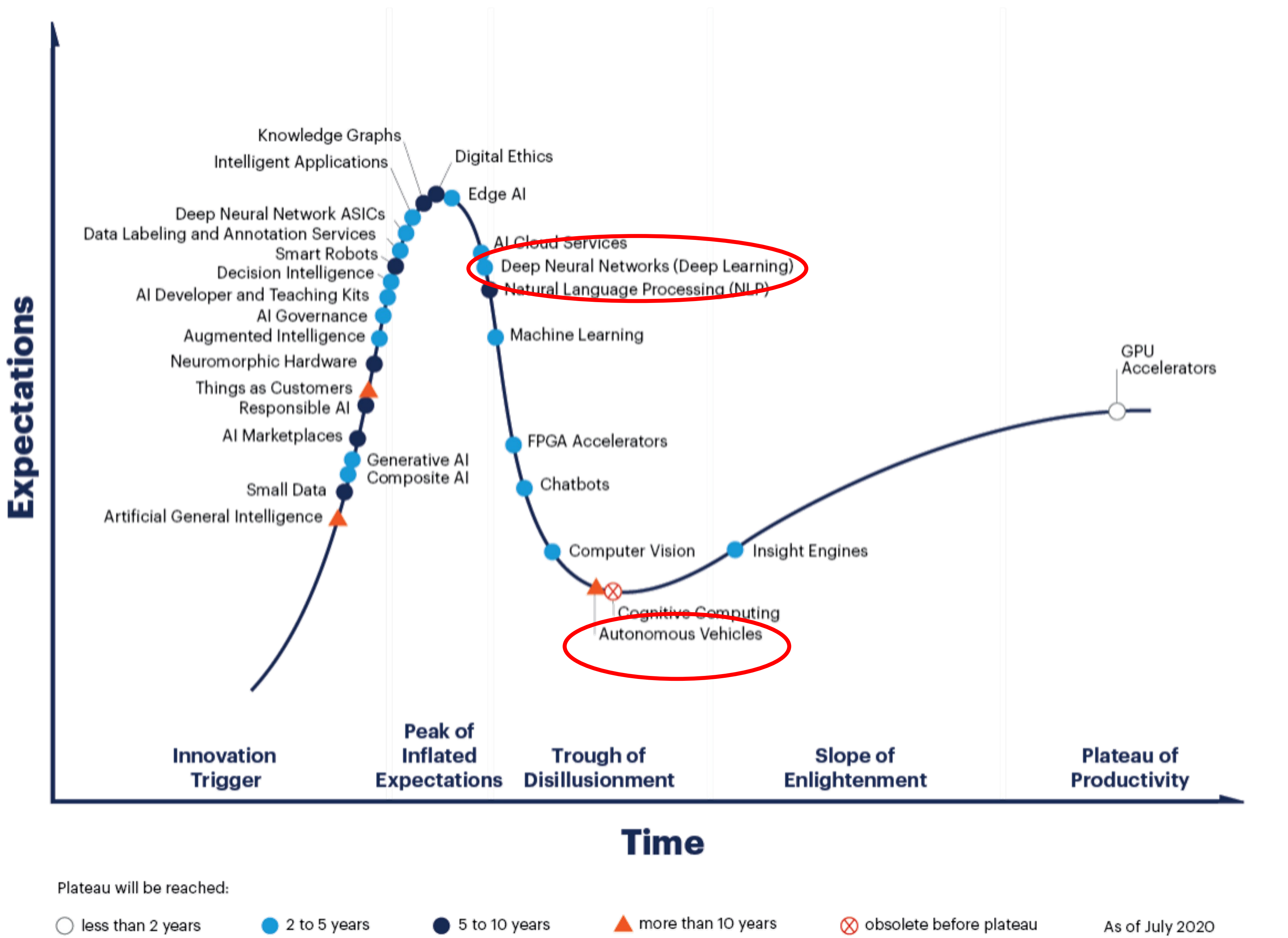
- Lots of achievements in vision field
- Not a magic tool!
  - Lack of adaptability and generality compared to human-vision system
  - Not able to build general-intelligent machine

Limitations of Deep Learning



Source: Gartner Hype Cycle for AI, 2019

Limitations of Deep Learning

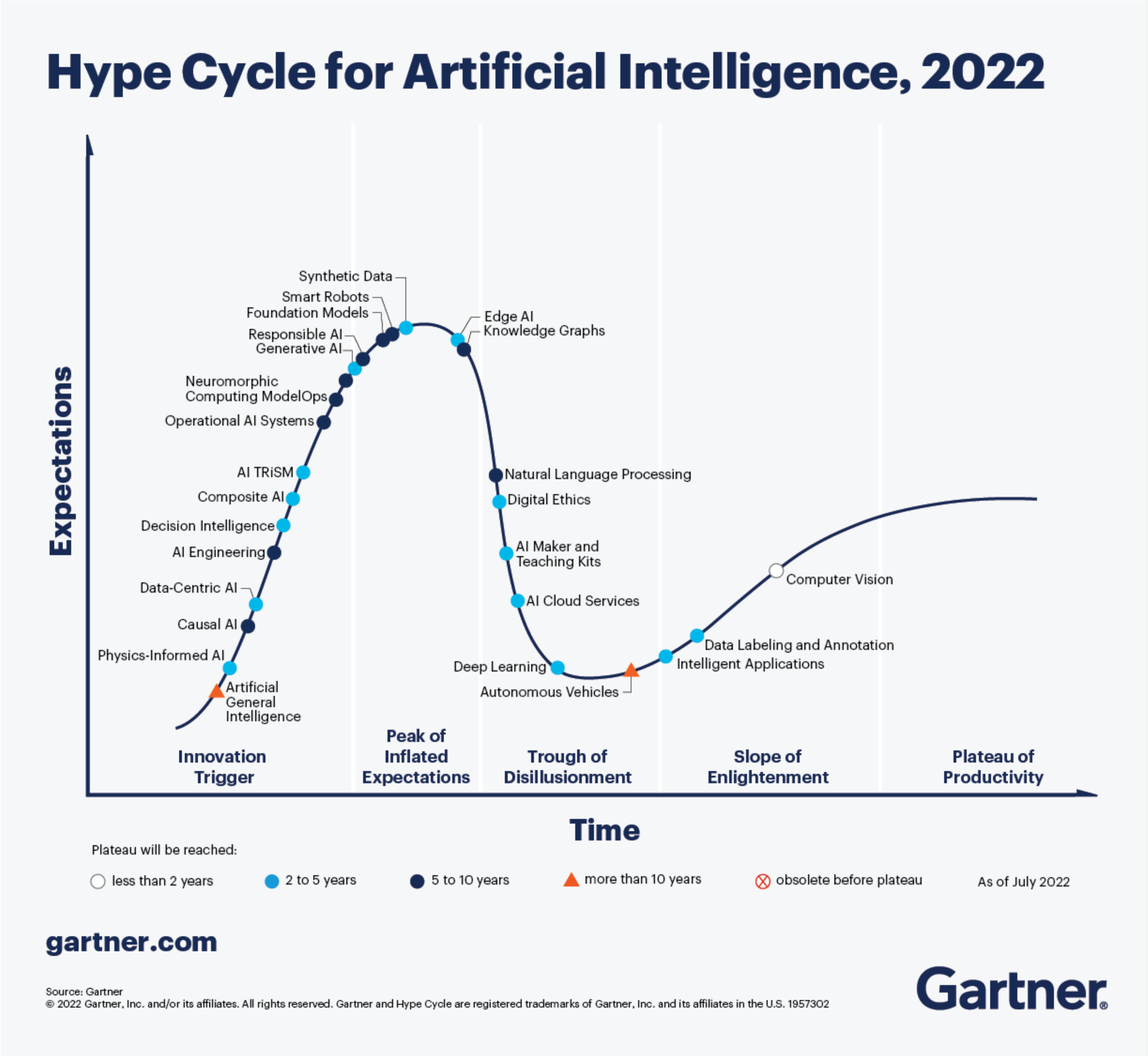


Source: Gartner Hype Cycle for AI, 2020

Limitations of Deep Learning

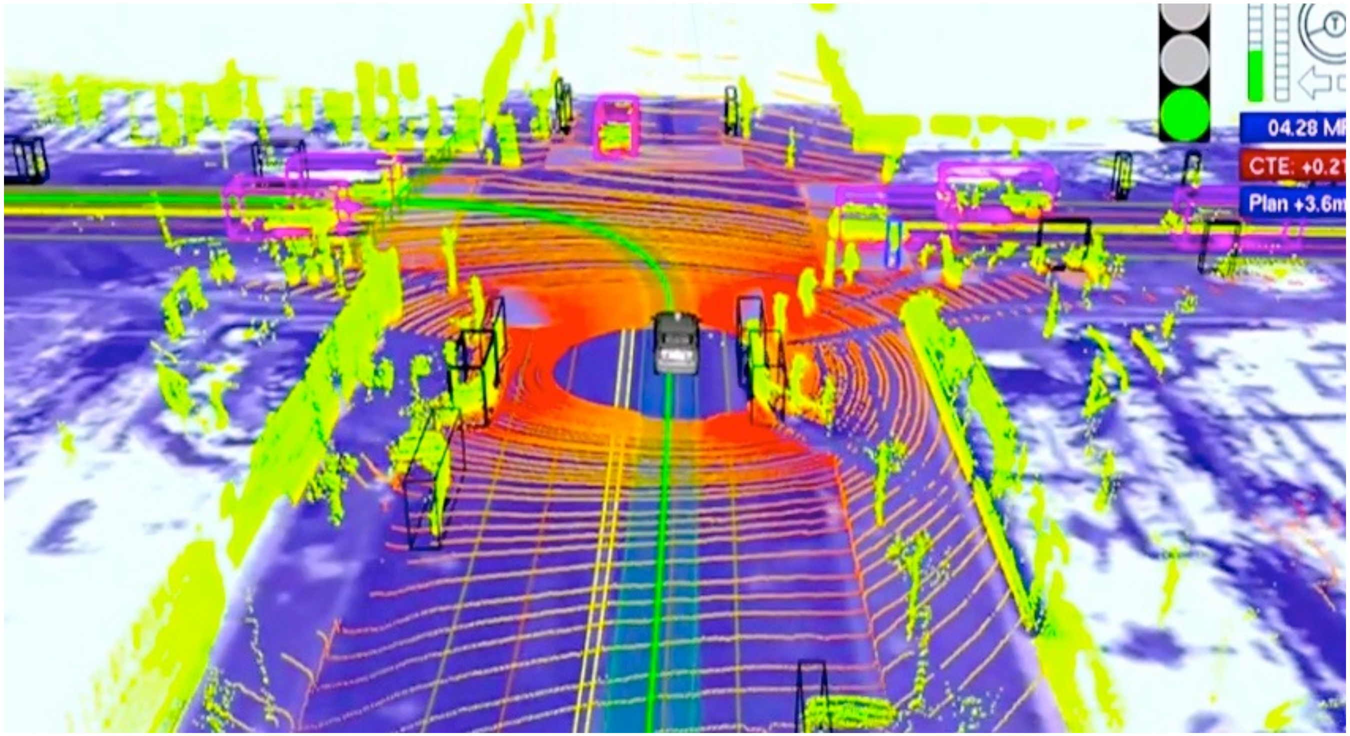
Source: Gartner Hype Cycle for AI, 2021

Limitations of Deep Learning



Limitations of Deep Learning

- Why cannot fit all real-world scenarios?



Source: Google

Source: Boston Dynamics

Limitations of Deep Learning

- Large amount of labeled data
  - Impressive achievements correspond to supervised learning
  - Expensive!
  - Sometimes experts & special equipment are needed

Limitations of Deep Learning

- Datasets may be biased
  - Deep Networks become biased against rare patterns
  - Serious consequences in some real-world applications (e.g., medical, automotive, … etc.)
  - Researchers should consider synthetic generation of data to mitigate the unbalanced representation of data

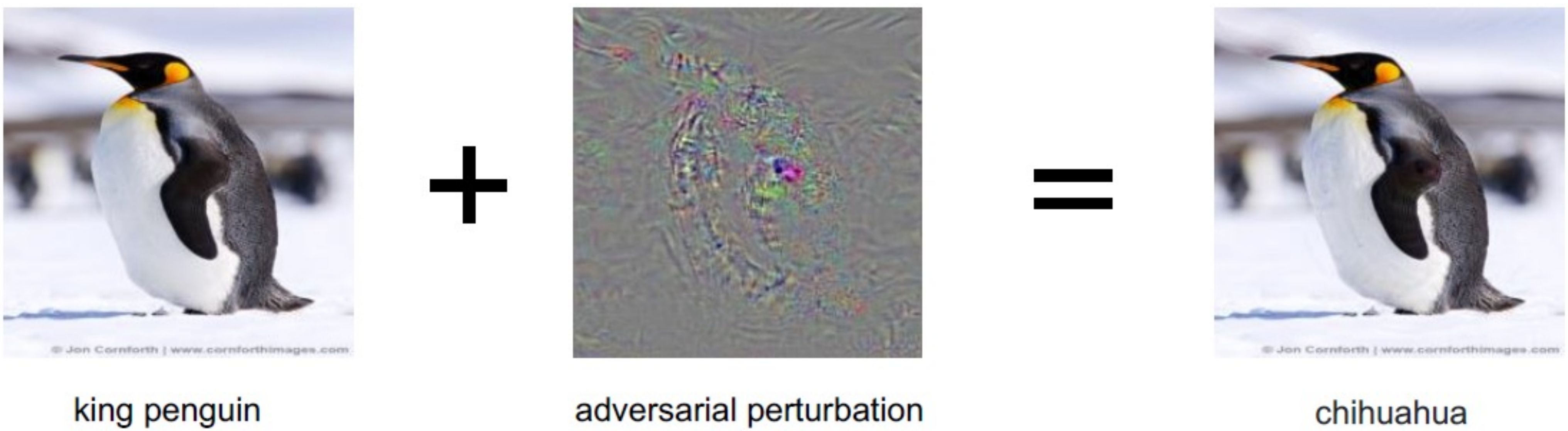
Limitations of Deep Learning

- Datasets may be biased

  - Classification may be sensitive to viewpoint
    - if one of the viewpoints is under-represented

Limitations of Deep Learning

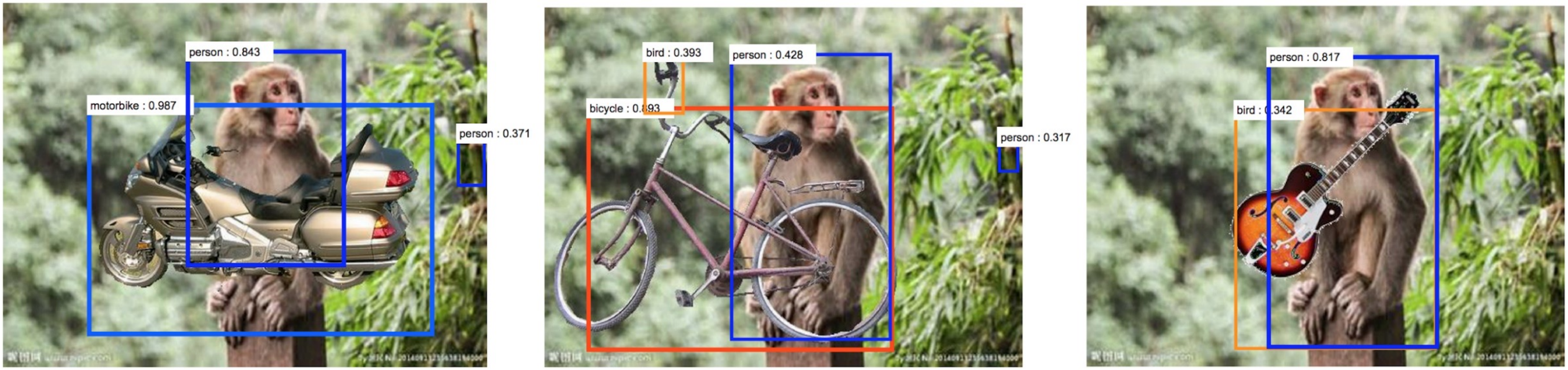
- Sensitive to standard adversarial attacks
  - Datasets are finite and just represent a fraction of all possible images



  - Add extra training, i.e., “adversarial training”

Limitations of Deep Learning

- Over-sensitive to changes in context
  - Limited number of contexts in dataset, i.e., monkey in jungle
  - Combinatorial Explosion!



Limitations of Deep Learning

- Combinatorial Explosion
  - Real world images are combinatorial large
  - Application dependent (e.g., medical imaging is an exception)
  - Considering compositionality may be a potential solution
  - Testing is challenging (consider worst case scenarios)

Limitations of Deep Learning

- Visual understanding is tricky
  - Mirrors
  - Sparse Information
  - Physics
  - Humor

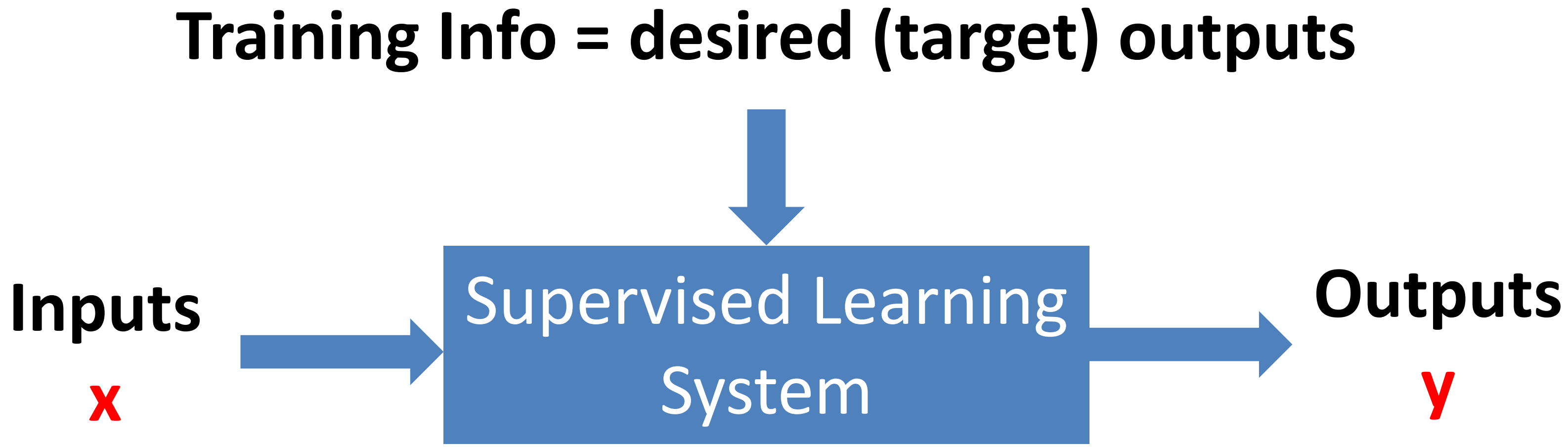
- Unintended results from fitness functions

Main Types of Learning

- Supervised Learning
- Unsupervised Learning
- Semi-Supervised Learning
- Reinforcement Learning

Supervised Learning

- Depends on labeled examples
- Learn the unknown function that corresponds outputs to inputs

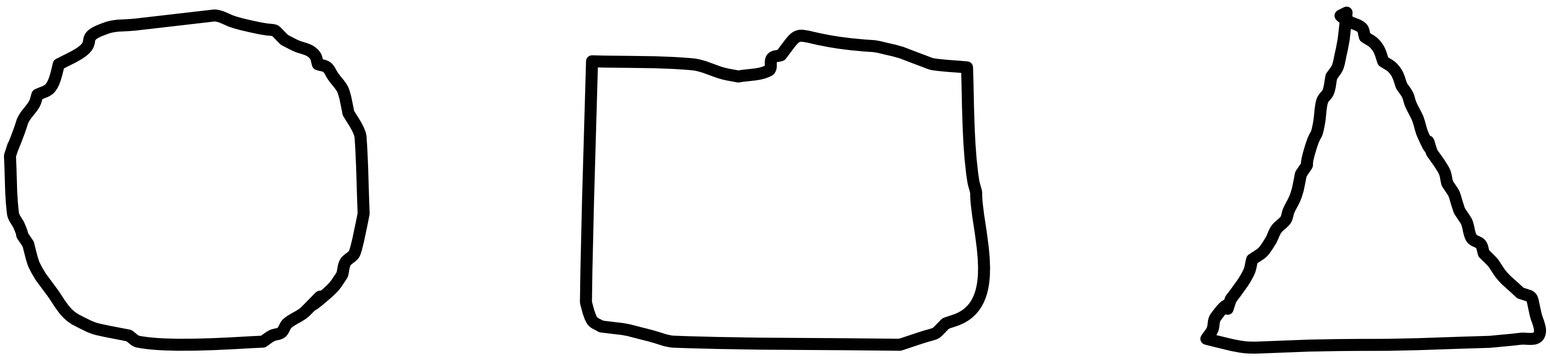


- Error = f(target output – actual output)
- Modeling $p(y|x)$

Example: Classify hand-drawn diagrams

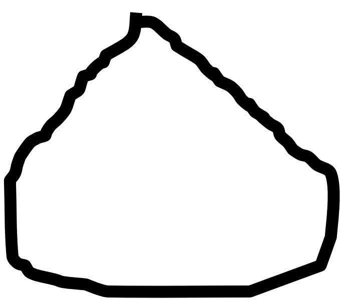
- Given a set of labeled training examples

- Classify any hand-drawn figures correctly!



circle rectangle triangle

- Complex problem!



Example: Classify hand-drawn diagrams

Training Phase

- Raw Data:
  - Training examples (labeled images)

Example: Classify hand-drawn diagrams

Training Phase

- Pre-processing:
  - Prepare data for processing (e.g., crop images, convert to grayscale instead of RGB … etc)

Example: Classify hand-drawn diagrams

Training Phase

- Features Extraction:
  - Design and select the key features to distinguish between shapes (e.g., ratio of area of figure to area of min enclosing circle)

Example: Classify hand-drawn diagrams

Training Phase

- Training a Classifier:
  - Machine learning algorithm
  - Choose classifier and training method
  - Examples: Bayes classifiers, artificial neural networks (ANNs), Support Vector Machines (SVM) … etc

Example: Classify hand-drawn diagrams

Training Phase

- Classifier:
  - Input: image
  - Output: label

Example: Classify hand-drawn diagrams

Training Phase

Supervised Learning!

Testing Phase

Unsupervised Learning

- Objective: Modeling p(x)
  - Learn generative model of the input data that tries to describe possible patterns and features of these inputs without the need for labeled outputs.

- Types:
  - Clustering problems
  - Association problems

Clustering

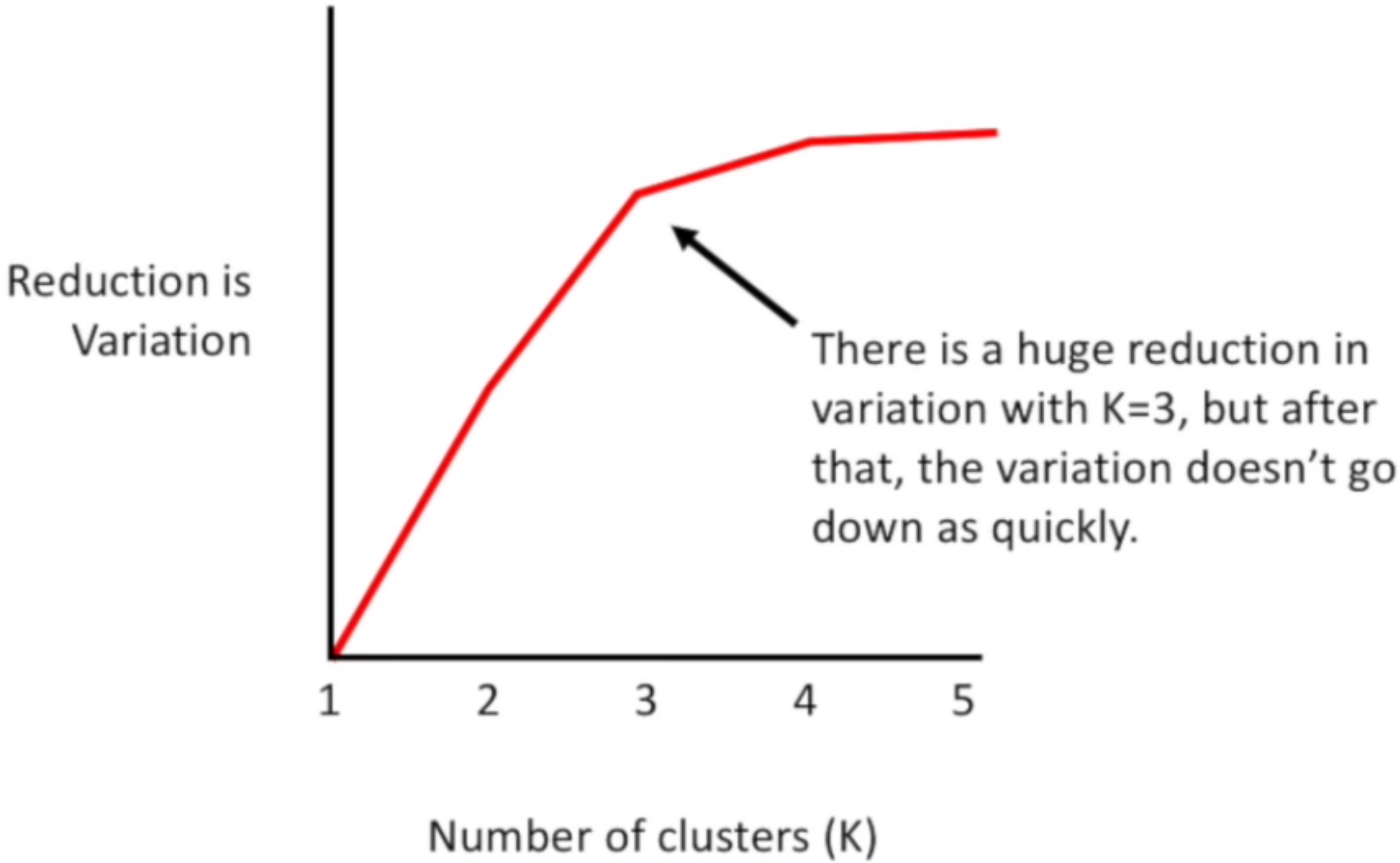
- Given unlabeled data
- Group data points of alike features together in some clusters
- Example clustering algorithms:
  - K-means
  - KNN clustering
  - Hierarchical clustering

K-means Example

- Steps:
  1. Select **k** points as centers of **k** clusters
  2. Cluster remaining points w.r.t. nearest center
  3. Calculate the mean of each cluster
  4. Repeat clustering based on means until no change
  5. Compute sum of variations of each cluster, i.e., sum of squared differences
  6. Repeat several times starting from step **1**
  7. Consider clustering centers with min variations as best option for **k** clusters
  8. Repeat the above algorithm staring from **k = 1** and keep increasing **k**
  9. Plot an elbow plot of the reduction in variations per **k**
  10. Select the elbow of the plot as your **K**

Check the example drawn on the white-board!

K-means Example



Source: [StatQuest, Josh Starmer]

Association

- Find dependencies within data
  - e.g., pasta and tomato-sauce sales in supermarket
- Generate dependency rules
  - e.g., pasta => tomato-sauce
  - customers who buy pasta are likely to buy tomato-sauce
- Goal: draw conclusions for better decisions
  - e.g., items placement in supermarket
- Example association algorithms:
  - Apriori
  - Frequent Pattern (FP) Growth

Semi-Supervised Learning

- Few labeled examples, i.e., sample data
- Lots of unlabeled data (cheap to obtain)
- Mixture of supervised and unsupervised techniques
- Makes use of unlabeled data by the help of labeled data

Semi-Supervised Learning

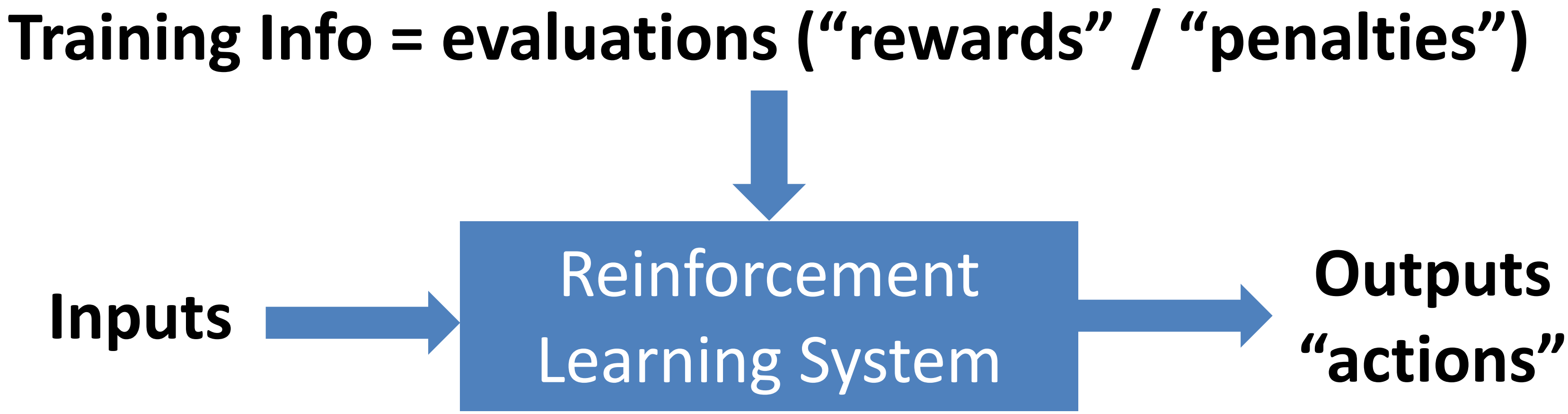
- Combining both types of data aims at improving performance
  - compared to applying supervised learning on the sample data

- Example applications:
  - Fraud detection in fintech, since num of know crimes is limited
  - Webpages mining

Semi-Supervised Learning Example Algorithm

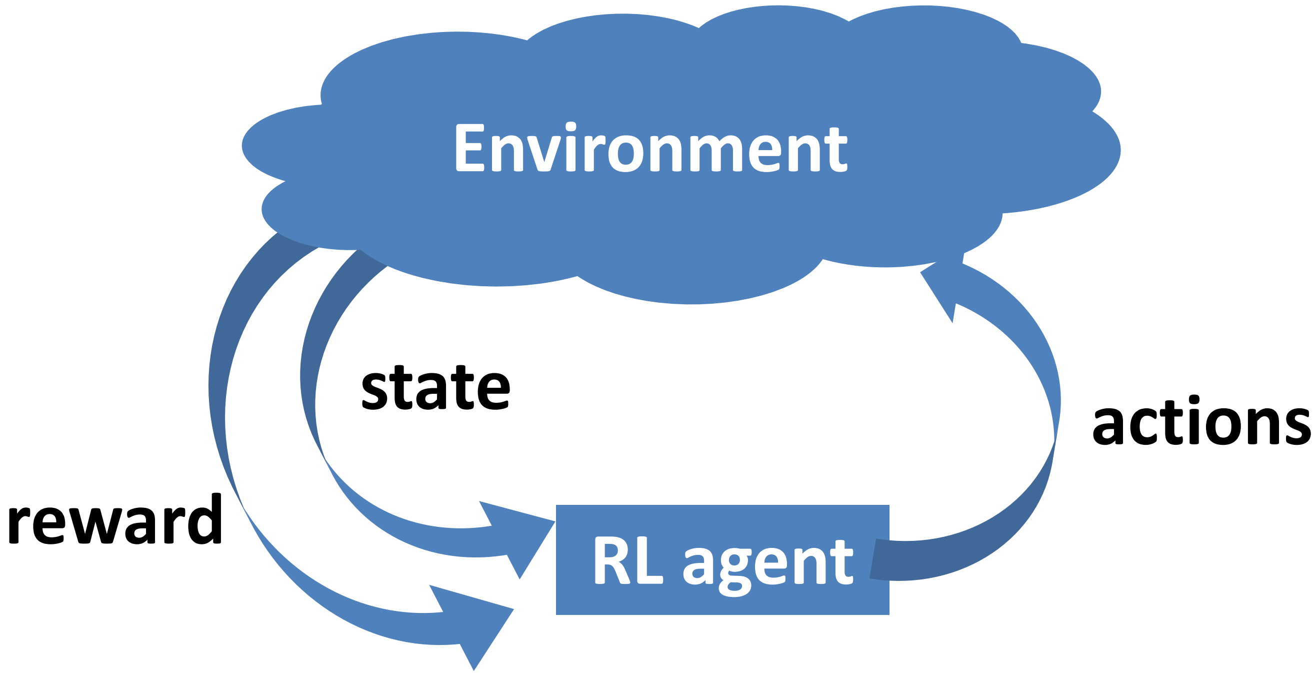
- Steps:
  1. Build supervised learning model based on labeled data, i.e., NN or Naïve Bayes … etc.
  2. Expectation step: Label unlabeled data using that model
  3. Maximization step: Retrain the model based on all the data
  4. Repeat starting from step 2 until convergence

Reinforcement Learning



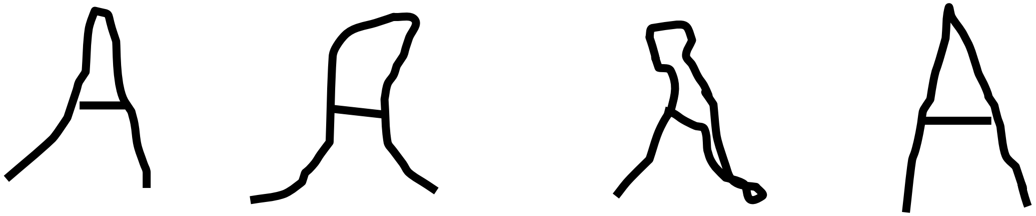
- Objective: Get as much reward as possible

Reinforcement Learning



What is a pattern?

- A pattern is an abstract object, such as a set of measurements describing a physical object
- Examples:
  - Hand-written characters



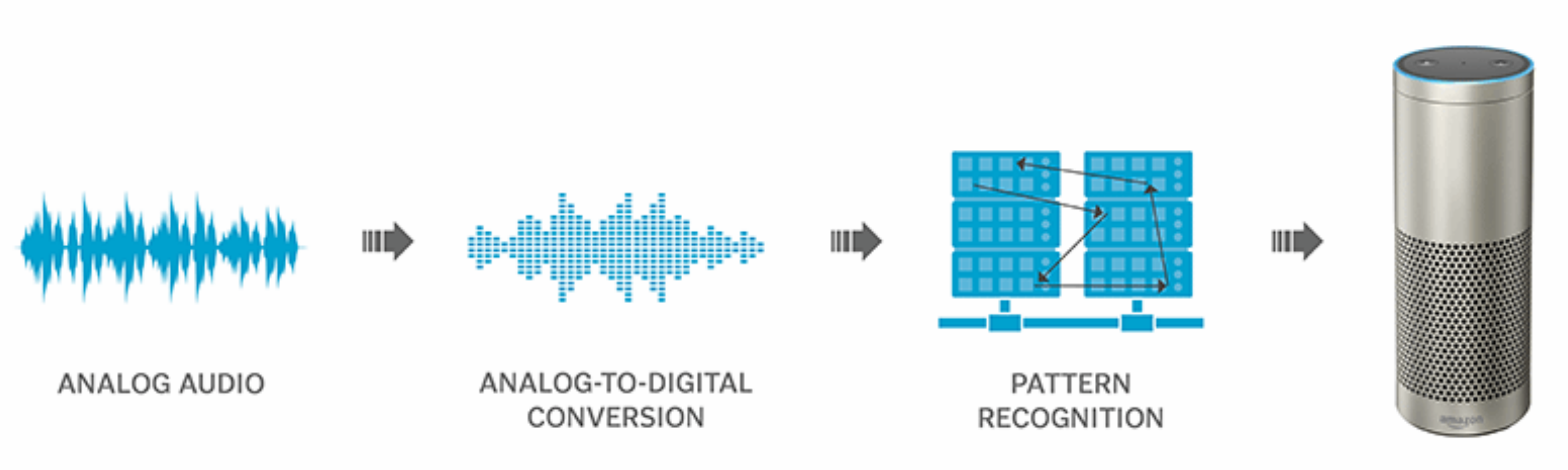
  - Finger prints

Pattern Classification

- *aka* pattern recognition
- Build intelligent machines to recognize patterns
- Examples:
  - recognize spoken words
  - recognize object in an image

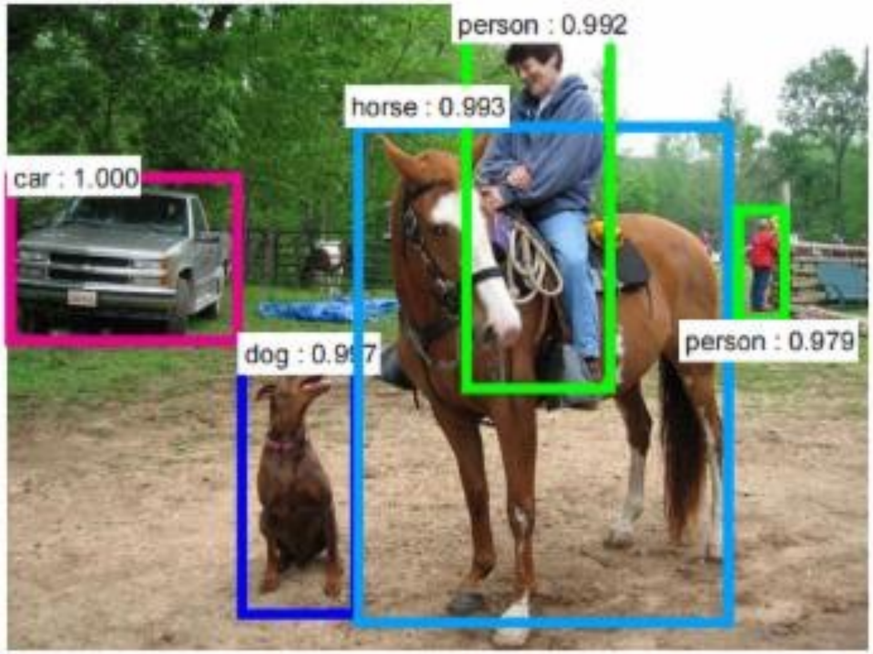
Speech Recognition

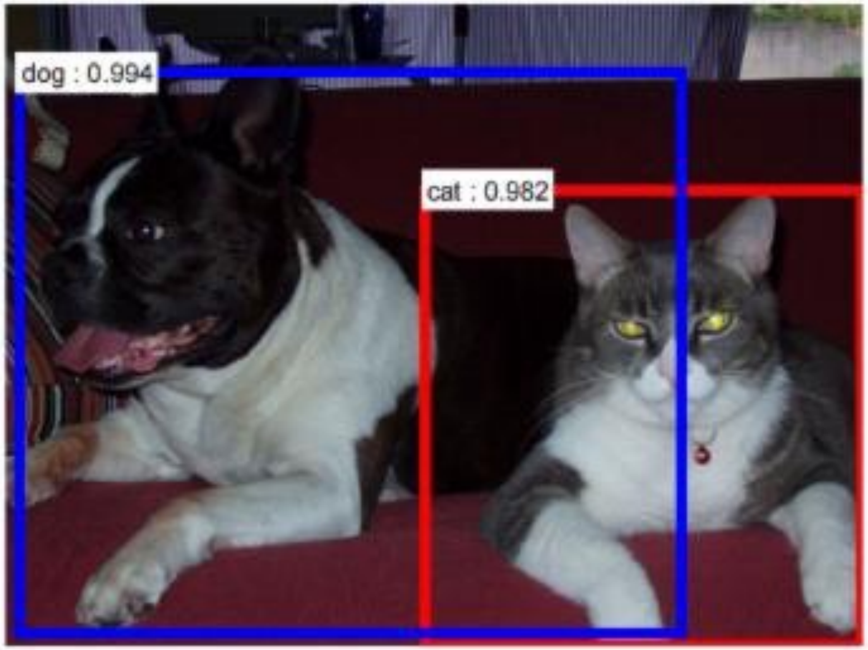
- Example:
  - recognize spoken words



Object Classification

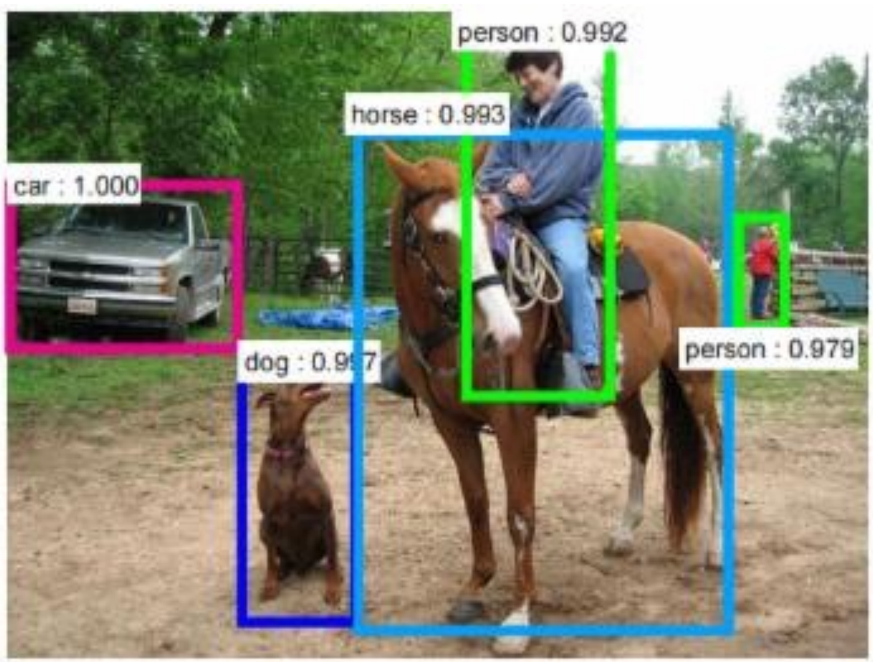
- Example: recognize object in an image:
  - Object recognition means that we implement an algorithm that recognizes objects in an image

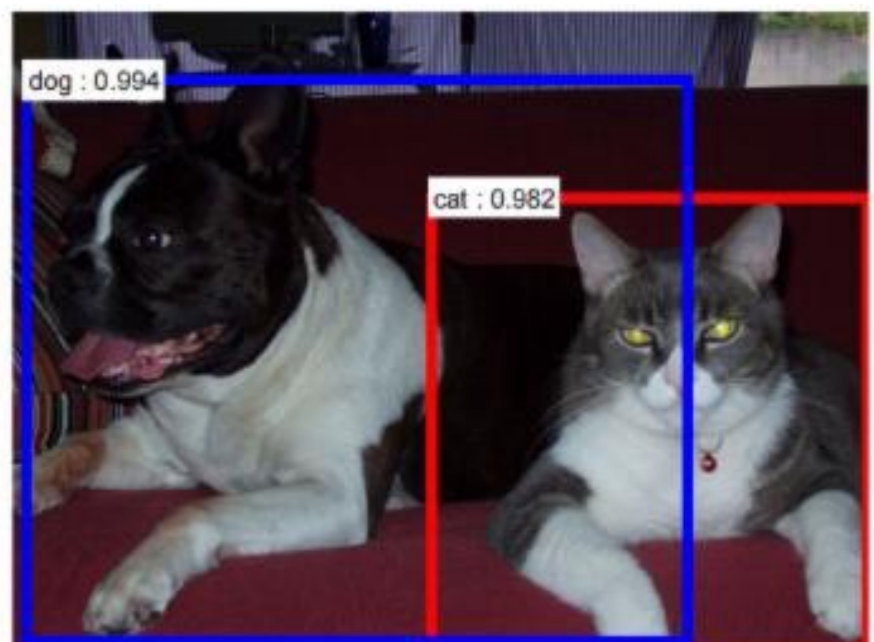




Object Classification

- Example: recognize object in an image:
  - Classify the objects: dog → class 1
person → class 2
car → class 3 … etc
  - Find features that are characteristic of each class





Hard Problem

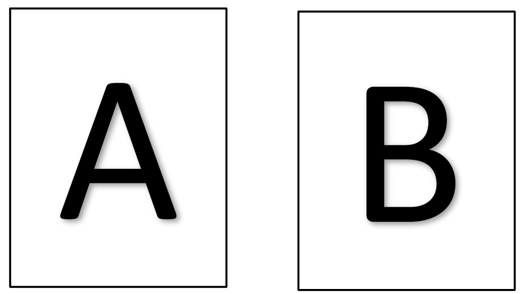
- Object recognition is a fairly hard problem due to variability of objects
  - Same objects can look different (e.g., rotated, different size, different angle of camera … etc)



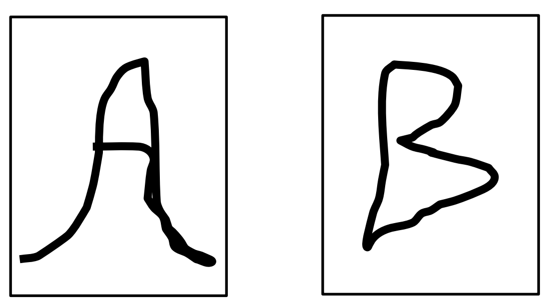


Optical Character Recognition (OCR)

- printed → accuracy 100%



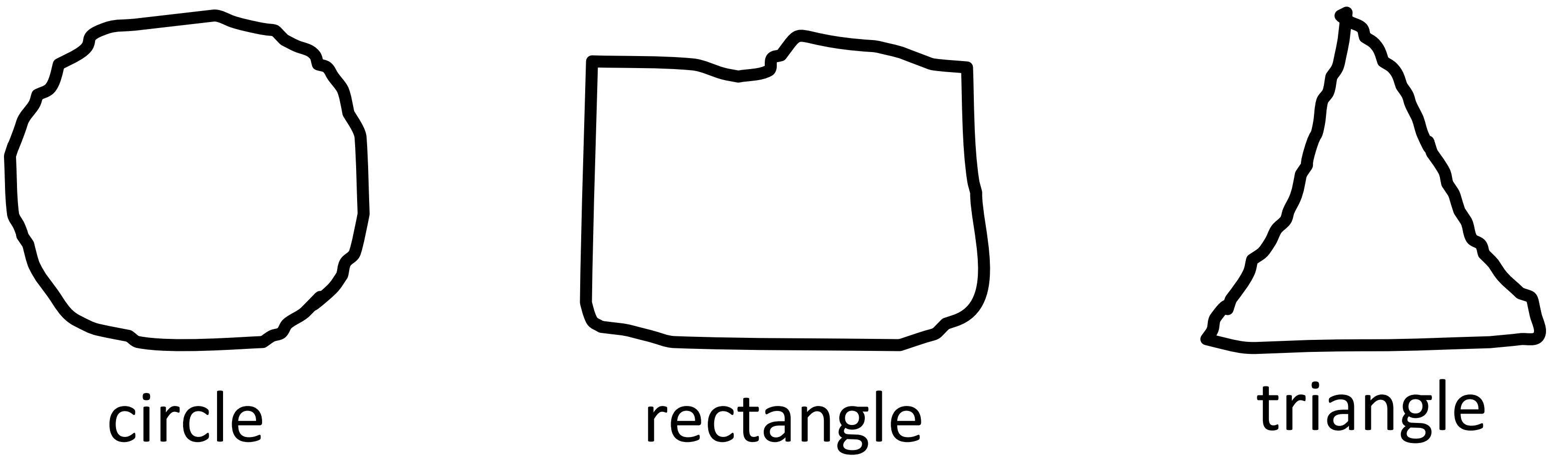
- hand-written → accuracy ≈ 99% to 99.5%



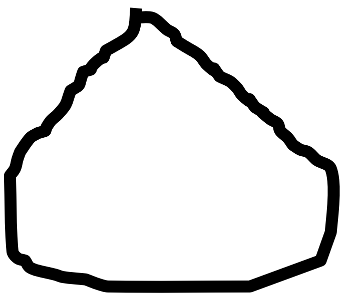
Example: Classify hand-drawn diagrams

- Given a set of labeled training examples

- Classify any hand-drawn figures correctly!



- Complex problem!



Example: Classify hand-drawn diagrams

Training Phase

Supervised Learning!

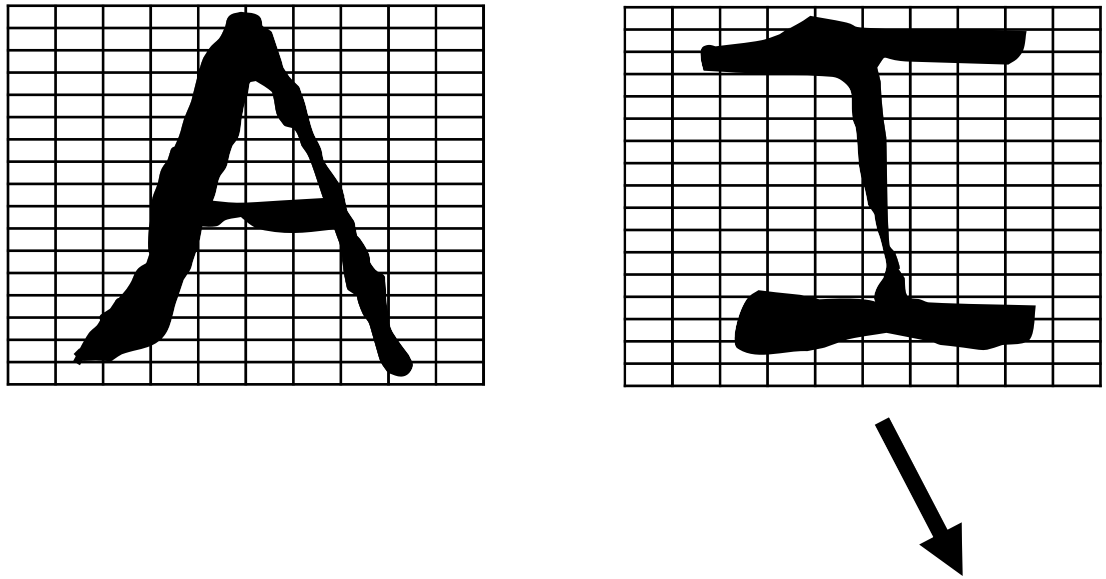
Testing Phase

Classification vs Regression

- Both problems tend to learn an unknown function that maps outputs to inputs → **function approximation**
- Function approximation is achieved via training
- Classification: predict a label
- Regression: predict a quantity (e.g., time series prediction, stock-market forecasting)

Example: OCR

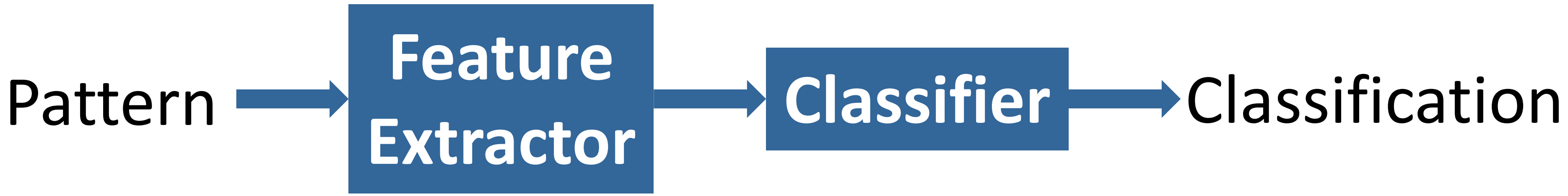
- Classify written text
- Images are a matrix of pixels (0's & 1's)



$$\begin{bmatrix} 0 & \cdots & 1\;1\;1 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & \cdots & & 0 \end{bmatrix}$$

Example: OCR

- Pattern classification system:



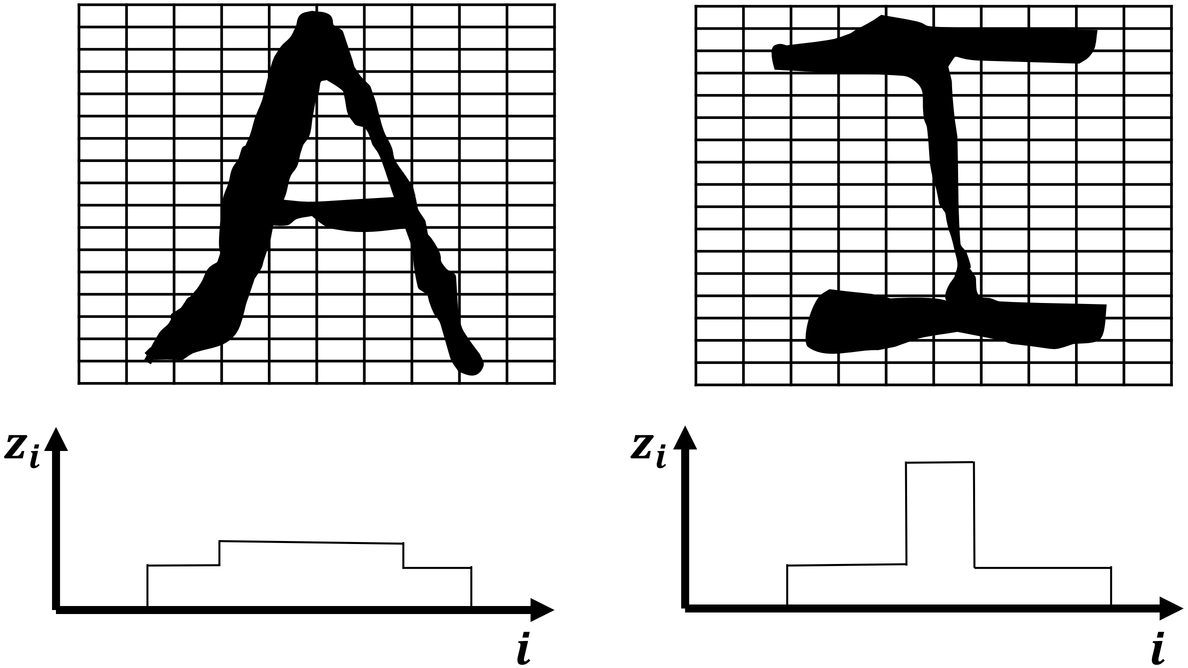
- Training set: collect large number (1000's) of example characters (different fonts, sizes and other variants)

Example: OCR

- Use the pixel matrix (or the pattern)
- Extract through an equation or an algorithm a set of features
- These features must be the most effective in discriminating between the different classes (characters in this example)

Example: OCR

- z_i = sum of black pixels along column i



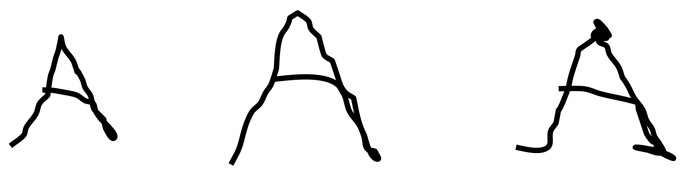
- Feature: $\mathbf{z} = max_i(z_i)$
  - z = 3 for ‘A’
  - z = 10 for ‘I’

Example: OCR

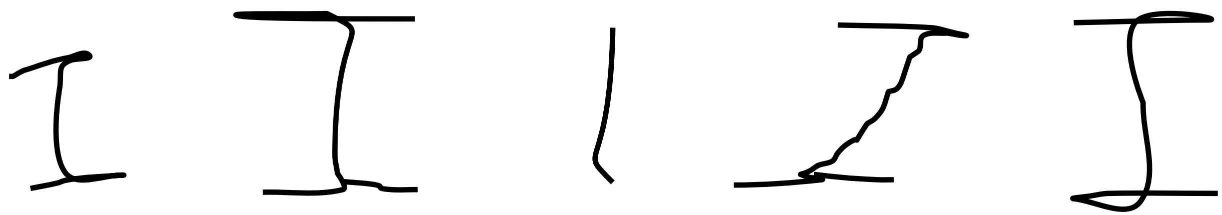
- **z** is our selected feature.
- It tends to be high in case of ‘I’, while it tends to be low for ‘A’
- It can effectively differentiate between these two classes

Example: OCR

- Construct a histogram based on the feature values of the training patterns



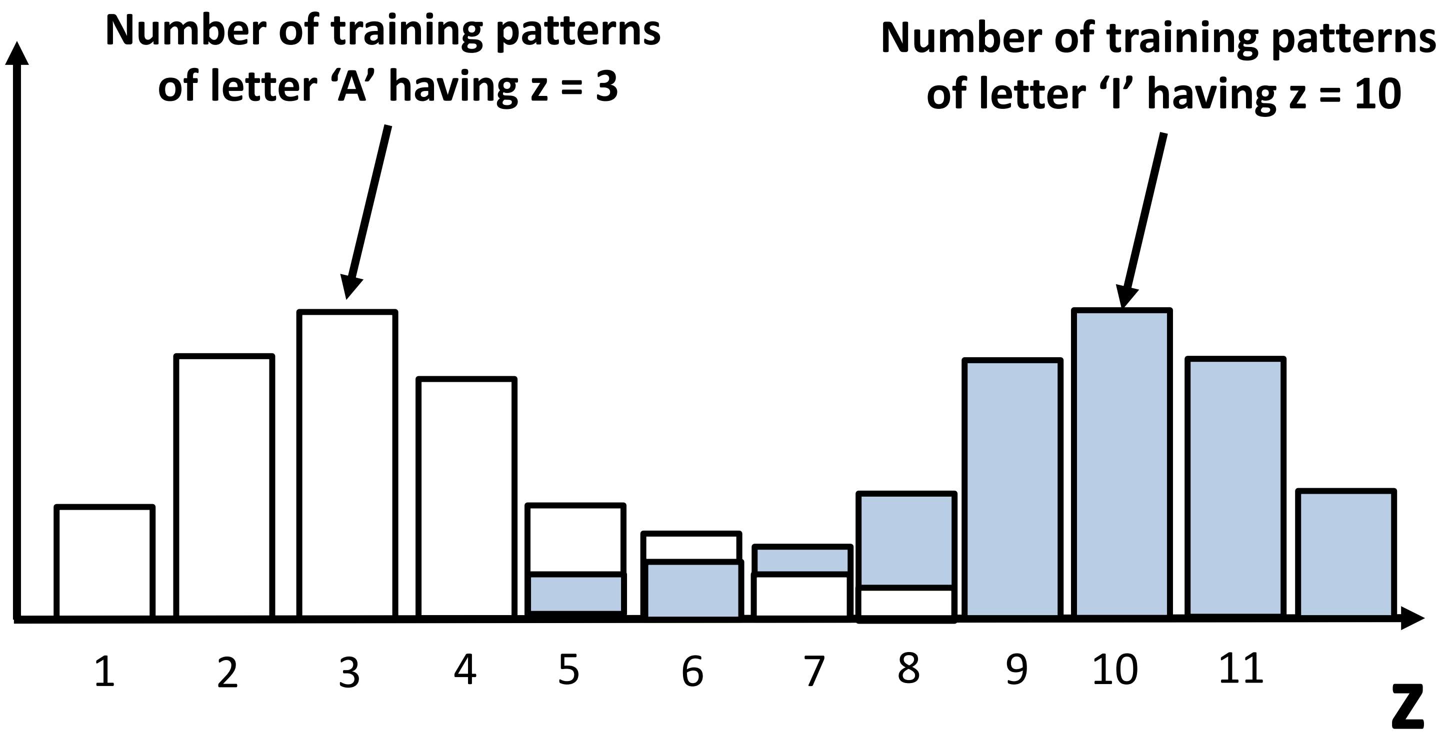
$z =$ **2** **3** **1** **4** **3**



$z =$ **8** **10** **9** **5** **7**

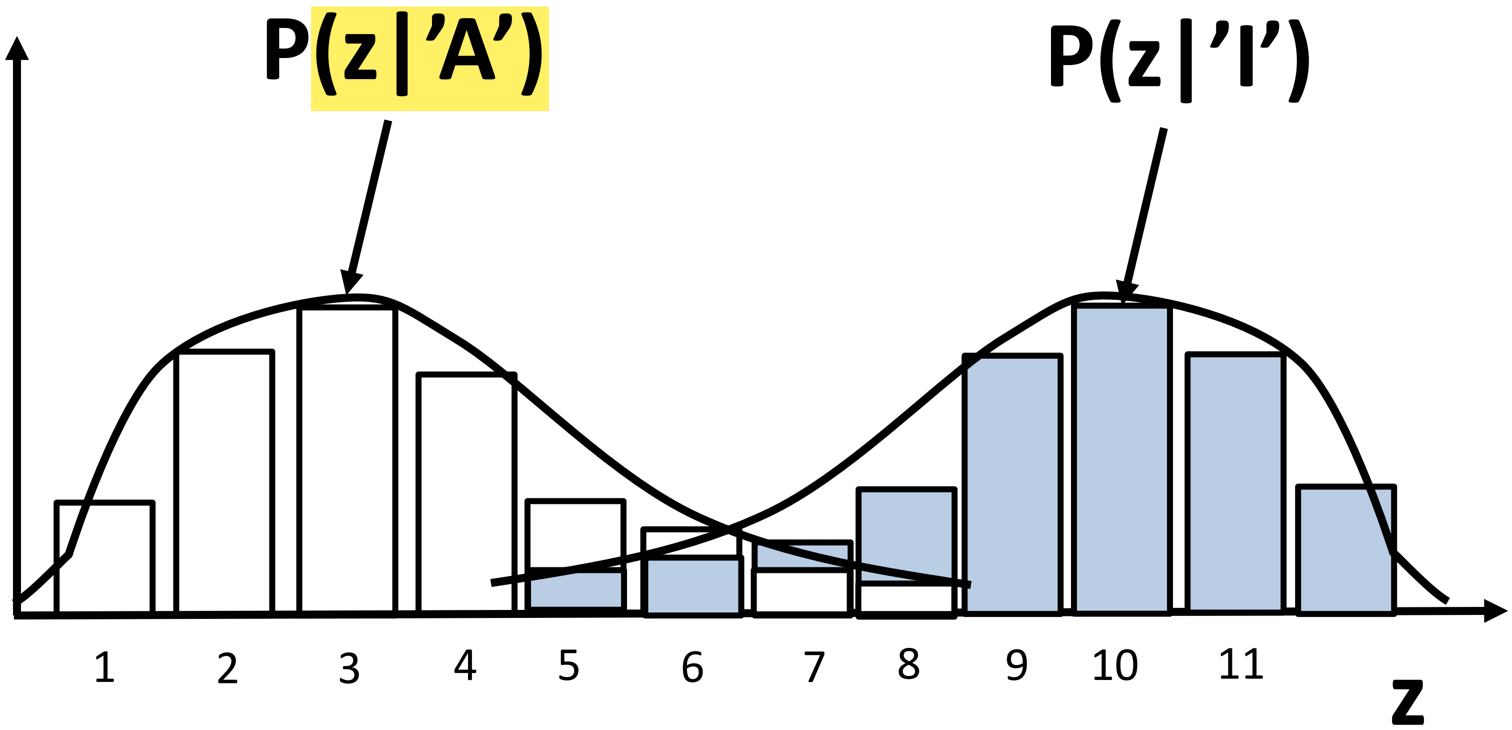
Example: OCR

- Construct a histogram based on the feature values of the training patterns



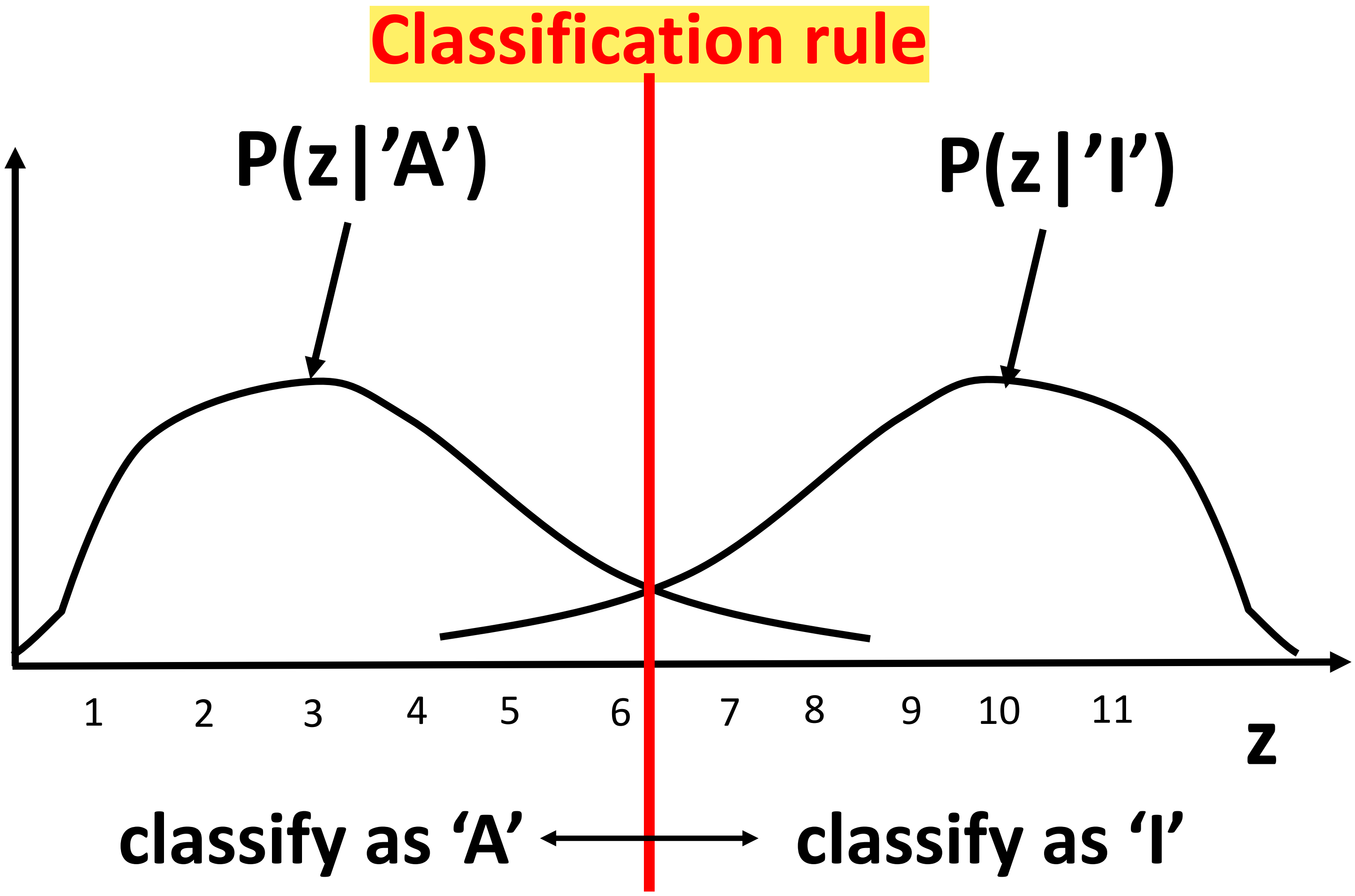
Example: OCR

- Estimate density functions



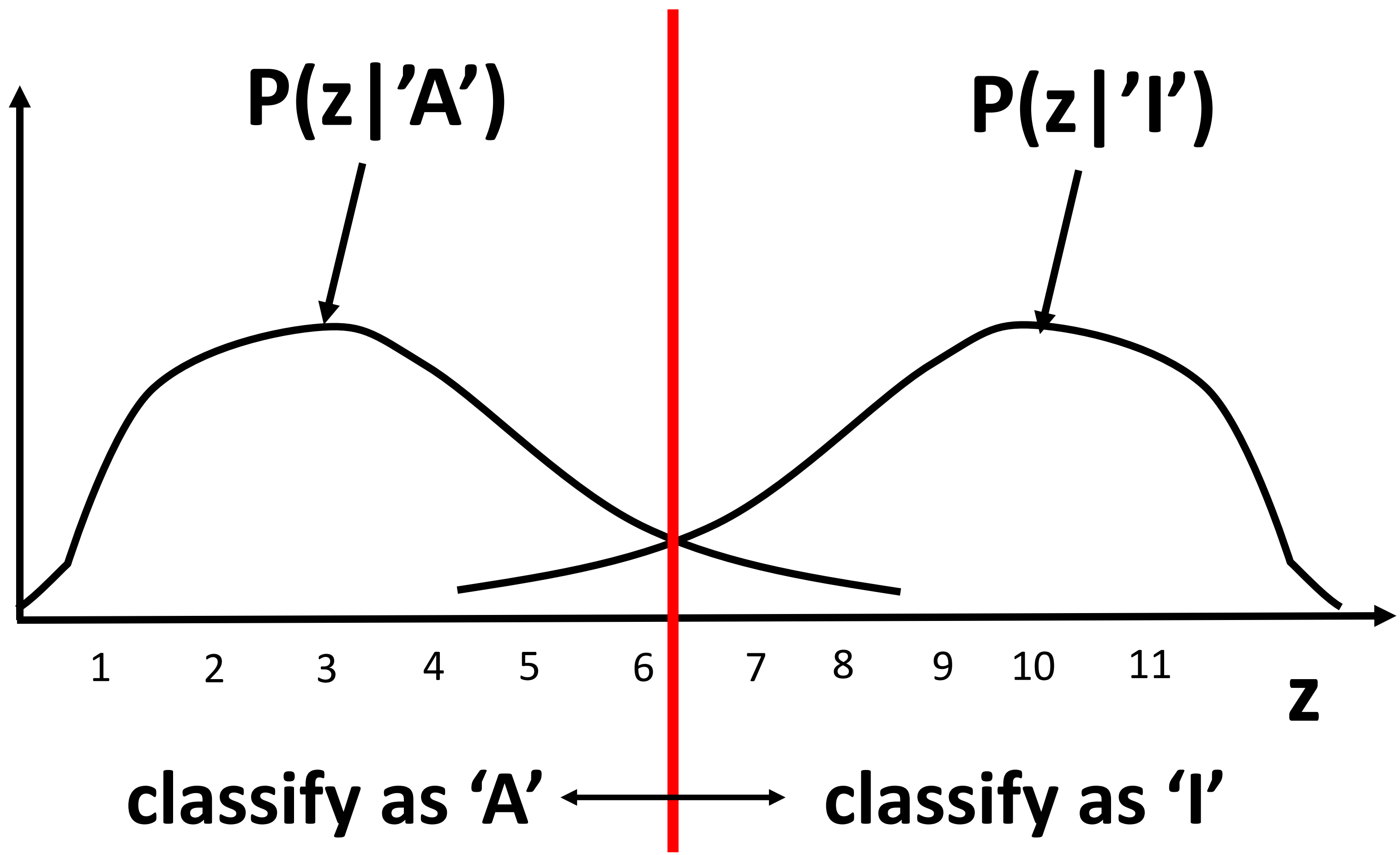
Example: OCR

- Classification rule



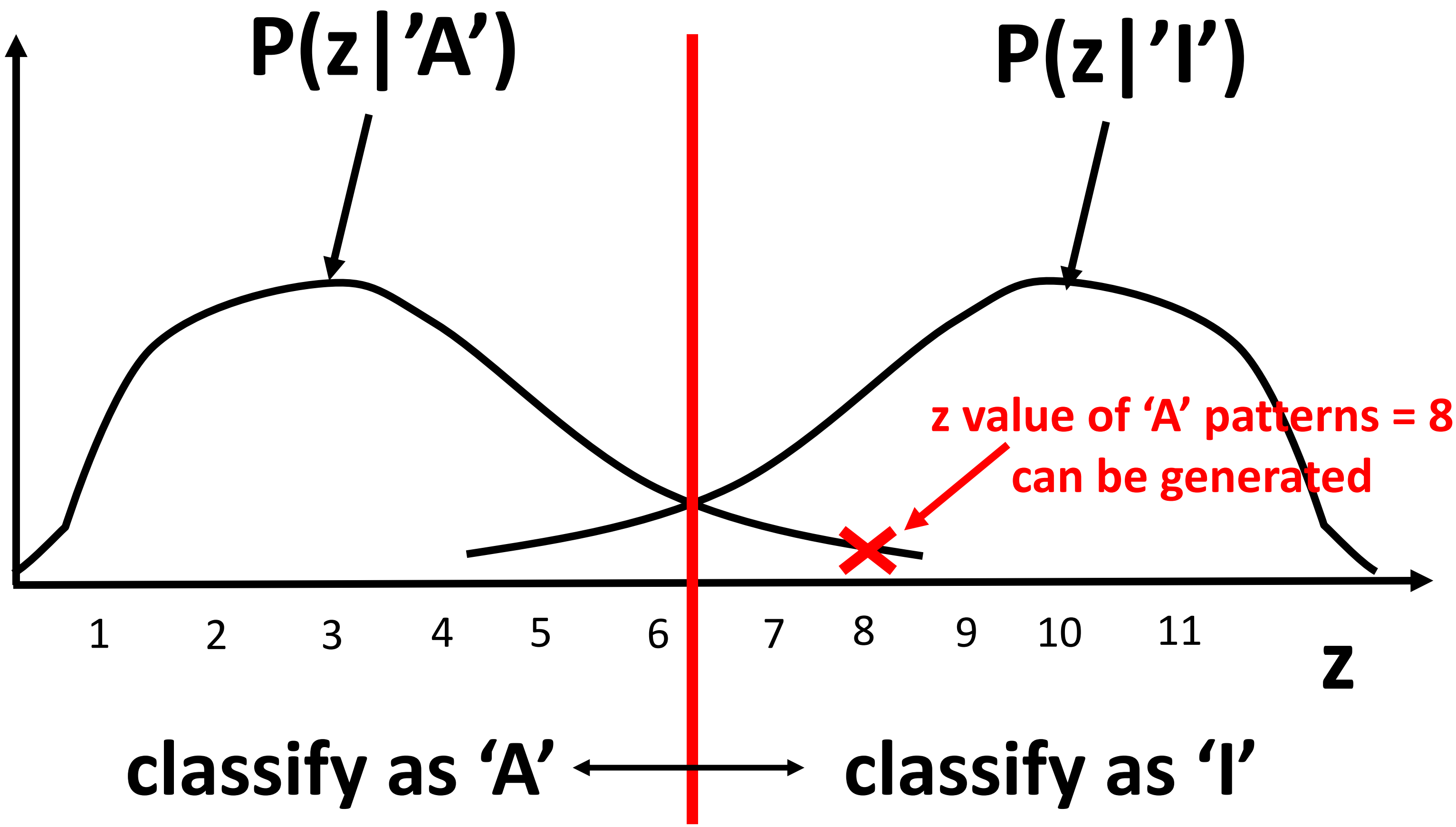
Example: OCR

- Some characters can produce feature values z that deviate some what from what is usually observed



Example: OCR

- For example one of the A's gave z = 8
- For such a pattern the classifier will produce a wrong answer (i.e., classification error)



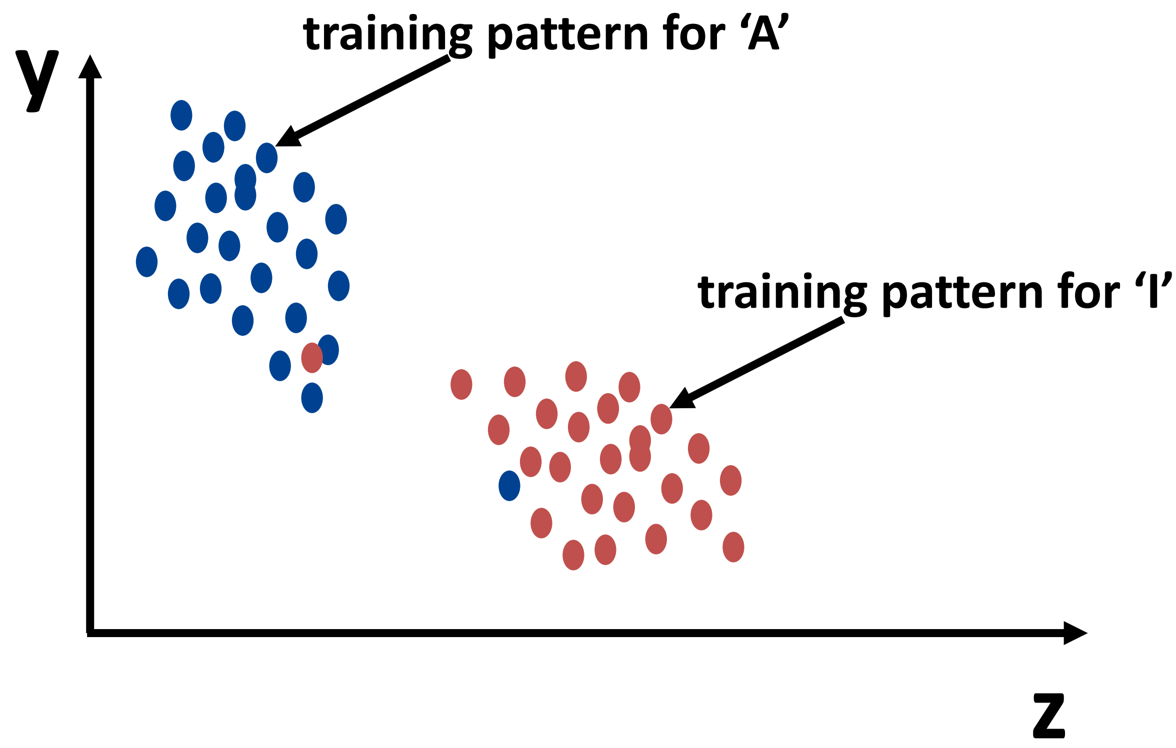
Example: OCR

- It is very difficult to design a classifier that yields no classification errors

- Classification errors occur due to the variability of the patterns typically encountered

Example: OCR

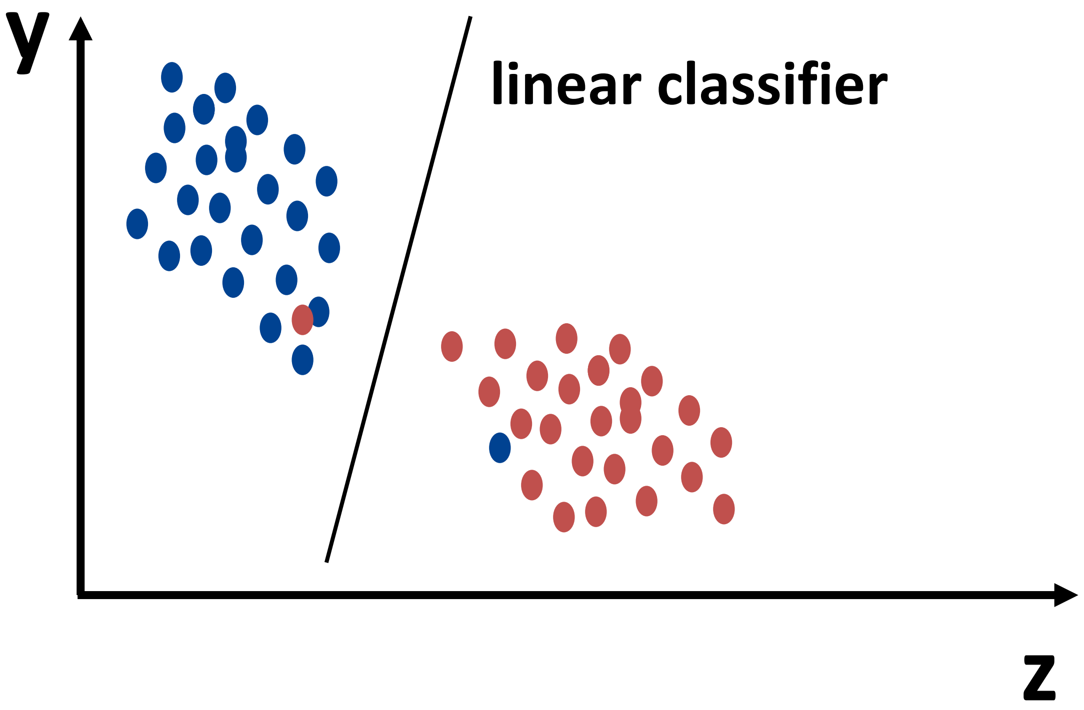
- Multiple features reduce the classification error



2d feature space for features y & z

Example: OCR

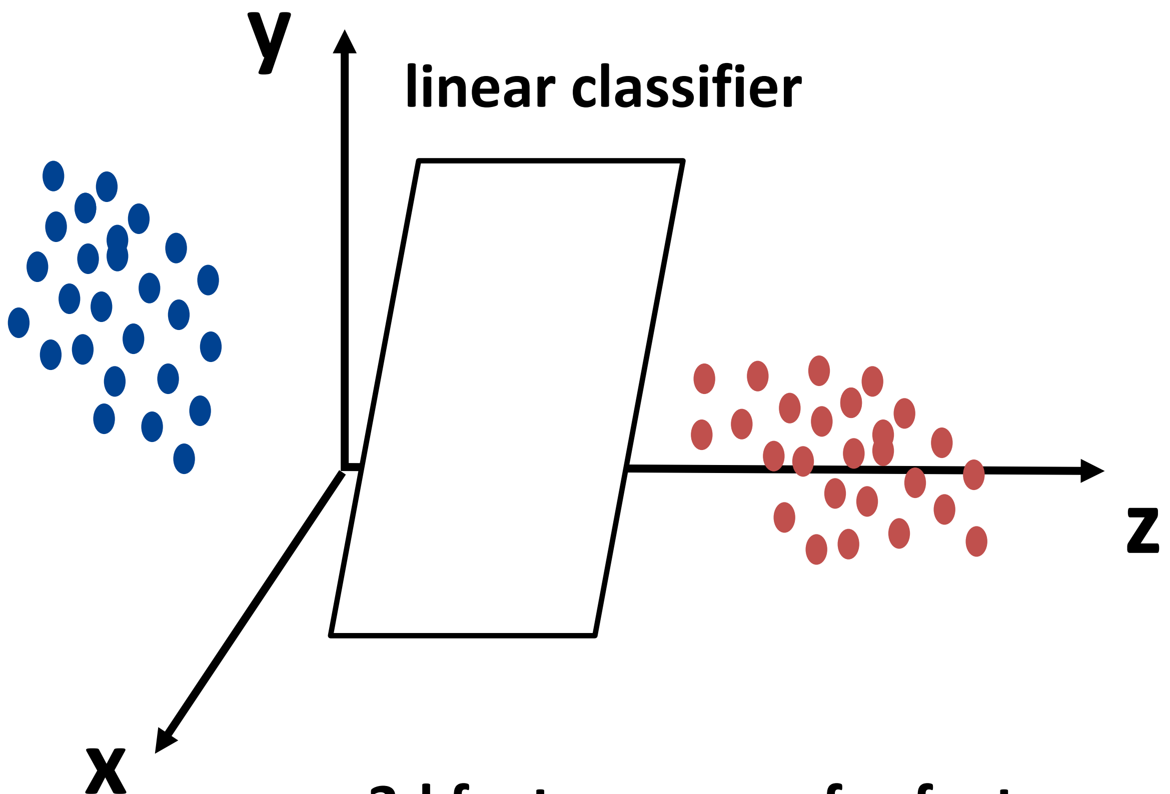
- Multiple features reduce the classification error



2d feature space for features y & z

Example: OCR

- Multiple features reduce the classification error



3d feature space for features x, y & z

Acknowledgment

- These slides have been created relying on previous materials by Amir Atiya, Stuart Russell, Peter Norvig, Bernhard Beckert, Scott Niekum, Selim Aksoy, Sargur Srihari & Sven Behnke